Improving Construction Site Inductions through the use of a Facial Emotion Recognition; a Study in Engagement

Charles Constant and Elizabeth Bates

Project Supervisor: James Ford

Civil, Environmental and Geomatic Engineering

31st March 2021

DECLARATION & PERSONAL STATEMENT

The project has been carried out by Charles Constant and Elizabeth Bates.

The authors have read and understood the College's policy regarding plagiarism and the submission of coursework. The authors confirm that, except for commonly understood ideas and concepts, or where specific reference is made to the work of other authors, the contents of this report are their own work. This dissertation is presented in 80 pages including references and appendices. It contains approximately 10700 words, 27 figures and 4 tables.

Statement by: Charles Constant

I estimate that my contributions to the parts of the project are as follows:

Report Sections:	%
Literature review	50
Theoretical development	50
Computational work	50
Experimental work	50
Data analysis	50
Report writing and production	50
Average	**50**

Signature: Date: 28.3.21

Statement by: Elizabeth Bates

I estimate that my contributions to the parts of the project are as follows:

Report Sections:	%
Literature review	50
Theoretical development	50
Computational work	50
Experimental work	50
Data analysis	50
Report writing and production	50
Average	**50**

Signature: E. Bates
ELIZABETH BATES

Date: 28.3.21

Abstract:

This project studies and uses facial emotion recognition methods to propose a novel means of quantify engagement in the context of construction site health and safety inductions.

Current perceptions of health and safety site inductions in the construction industry are studied. The relationship between affect and knowledge retention is studied with the aim of understanding how this applies to improved site safety. The term ‘engagement’ and the role of emotions within it are identified.

A preliminary method of quantifying engagement is proposed and trialled. The outputs of this method are compared to two independent markers of engagement: an affect rating scale (Positive Affect Negative Affect Schedule) and a cognitive engagement survey. Suggestions are made for the way in which the method can be further validated, improved and industry use cases are identified.

Table of Contents

1.0 Introduction and Scope of the Project 9

2.0 Background and Literature Review 10

2.1 Broader Context of H&S.. 10

2.2 Site Inductions and Safety Culture 11

2.3 General Engagement 12

2.4 The Role Emotions Play in Engagement... 14

2.5 Quantifying Emotions 17

2.5.1 Subjective and Qualitative Methods.. 18

2.5.2 Objective and Quantitative Methods 19

2.6 Computer vision and its applications to emotion recognition................... 20

2.7 The Theory Behind Emotion Recognition in Machine Learning 24

2.7.1 Computer Vision 24

2.7.2 Convolutions 26

2.7.3 CNNs as an Evolution of ML.. 28

3.0 Work, Testing and Results 29

3.1 Project Specific Engagement Metric 29

3.2 Methodology.. 32

3.2.1 Initial Methodology 33

3.2.2 Background on DeepFace 34

3.2.3 Data Collection and Exploratory Data Analysis................................ 35

3.2.3.1 From Video to CNN Output... 35

3.2.3.2 From CNN Output to Practical Results 37

3.2.3.3 Exploratory Data Analysis and Quality Check 40

3.3 Measures of Engagement 43

3.3.1 Engagement Data 43

3.3.2 Quantifying Cognitive Engagement... 44

3.3.3 Quantifying Emotional Engagement Data 45

3.3.4 Numerical Contextualisation 45
3.3.5 Application example 1: Isolating engaging events 50
3.3.6 Application Example 2: Isolating key variables 51
3.4 Program Reflections 53
3.5 Limitations 53
4.0 Future Work 55
5.0 Conclusion 57
6.0 GitHub Code Repository Link https://github.com/CharlesPlusC/FER 58
7.0 References 58
8.0 Appendices 70
8.1 Appendix A – PANAS and Engagement Questionnaire Forms Used 70
8.2 Appendix B – PANAS and Engagement Questionnaire Results 72
8.3 Appendix C – Positive and Negative Emotion and Variance scores 81

Table of Figures

Figure 1 A flowchart showing the general approach to this project 10
Figure 2 Representation of general human and computer vision sequence (Susskind et al.,2007)... 25
Figure 3 Visual representation of an image to matrix of pixels and numeric values (Levin, 2020) 25
Figure 4 Graphical representation of the three colour matrices that compose a screen image.. 26
Figure 5 Structure of a perceptron unit... 26
Figure 6 The method by which a feature map is created................................. 27
Figure 7. Manually encoded feature extractors for images of a person (Felzenszwalb et al. 2008) 28
Figure 8 Illustration of how a CNNs multiple layers fit together (Mathworks,2020).. 28
Figure 9 First and second layer of filters of a CNN learning to classify images (left image) and an example of final layer features (three images on right). Taken from Lee et al.,(2015) 29
Figure 10 Flowchart showing the process of going from recorded sessions to final engagement scores 30
Figure 11 Program Input 32
Figure 12 Example output of an 'angry' emotion 33
Figure 13 Example output of a 'happy' emotion... 34
Figure 14. Application of DeepFace to an image with multiple faces present- as might be the case in a site induction. The percentage score displayed above the boxes represent the model's certainty that the faces are faces. (Image from Guardian News,2018) 35
Figure 15 Raw data plotted for video 1 of participant 1. 2 frames were processed for every second of video. 37
Figure 16. Emotion data for video 1 of participant 1. Positive and Negative emotions have been grouped. Note that the total percentage does not add up to 100% as a significant proportion of the recorded emotion is ‘neutral’.............. 38

Figure 17. Rolling average of positive and negative emotions for session 11 of participant 2 39
Figure 18. This graph presents the same information as the one above (Figure 12), with the addition of neutral. Note that neutral usually accounts for most of the emotion present during a session 39
Figure 19. Illustration of how data was taken from 'raw' CNN output to a format in which it could be analysed to identify engaging events during the recorded sessions 40
Figure 20 Emotion strength during a short test video. Different faces were pulled with the intention of triggering different outputs 41
Figure 21 Emotion strength for Video 1 of participant 2:Small Group; Q and A session; Some interaction with lecturer 42
Figure 22 Emotion strength for Video 2 of participant 2: Small Group; One interaction with lecturer around the midpoint of the lecture 43
Figure 23 Graphical representation of the program outputs against the relevant qualitative outputs for participant 1 throughout 8 sessions 48
Figure 24 Graphical representation of the program outputs against the relevant qualitative outputs for participant 2 throughout 8 sessions 49
Figure 25 Positive and negative emotion of the authors throughout a pre-recorded site induction. Arrows indicate where different changes in emotion are believed to be due to a change in the content. Cognitive and emotional engagement scores are presented below 51
Figure 26 Average difference in negative emotion grouped by presence of interactivity for all recordings of the authors 52
Figure 27 Average difference in variance grouped by presence of interactivity for all recordings of the authors 52
Figure 28 Positive and negative emotion, with corresponding variance score of all recordings of participant Figure 29 Positive and negative emotion, with corresponding variance score of all recordings of participant2 82

Table of Tables

Table 1 Comparison of quantitative methods to measure emotion.................... 22

Table 2 Example 'raw' output from the CNN for the first 5 frames of one session 37

Table 3 Spearman's correlation rank and t-test score table comparing the outputs of the CNN with independent markers of engagement 46

Table 4 The Cognitive Engagement Scores for each Session 78

1.0 Introduction and Scope of the Project

This project aims to work towards improving the current state of H&S outcomes on construction sites in the UK.

The project's objectives have been targeted at H&S site inductions.

The following objectives will be pursued in an attempt to advance towards the aim:

- Studying the role of site inductions on H&S outcomes
- Studying the role of participant engagement in improving inductions
- Understanding the role of emotion in engagement
- Comparing and selecting methods to quantify engagement
- Finding a method for the effective exploitation of the link between emotion and engagement to identify critical components of engaging inductions.
- Preliminary testing of the method, and identification of its limitations. Exploration of the methods applications, enabling further research to stem from it.
- Research and development of an engagement metric.

The initial aim of this work was to build on a method proposed by Ninaus et al. (2019) and apply an adapted version of it across construction sites in London. This would have enabled the analysis of multiple inductions, resulting in the identification of the key variables involved in successful site inductions.

In light of the current pandemic, attention has been redirected at making an effective method that future students and interested parties may use, with the intention to facilitate the pursuit of our initial aim. Additionally, focus will be turned to developing a metric which determines 'engagement' based on the outputs of the FER and accompanying questionnaires.

Finally, as a proof of concept, the method will be tested on a sample of videos. Limitations will be identified, and the method adjusted if needed. Figure 1 illustrates the project plan.

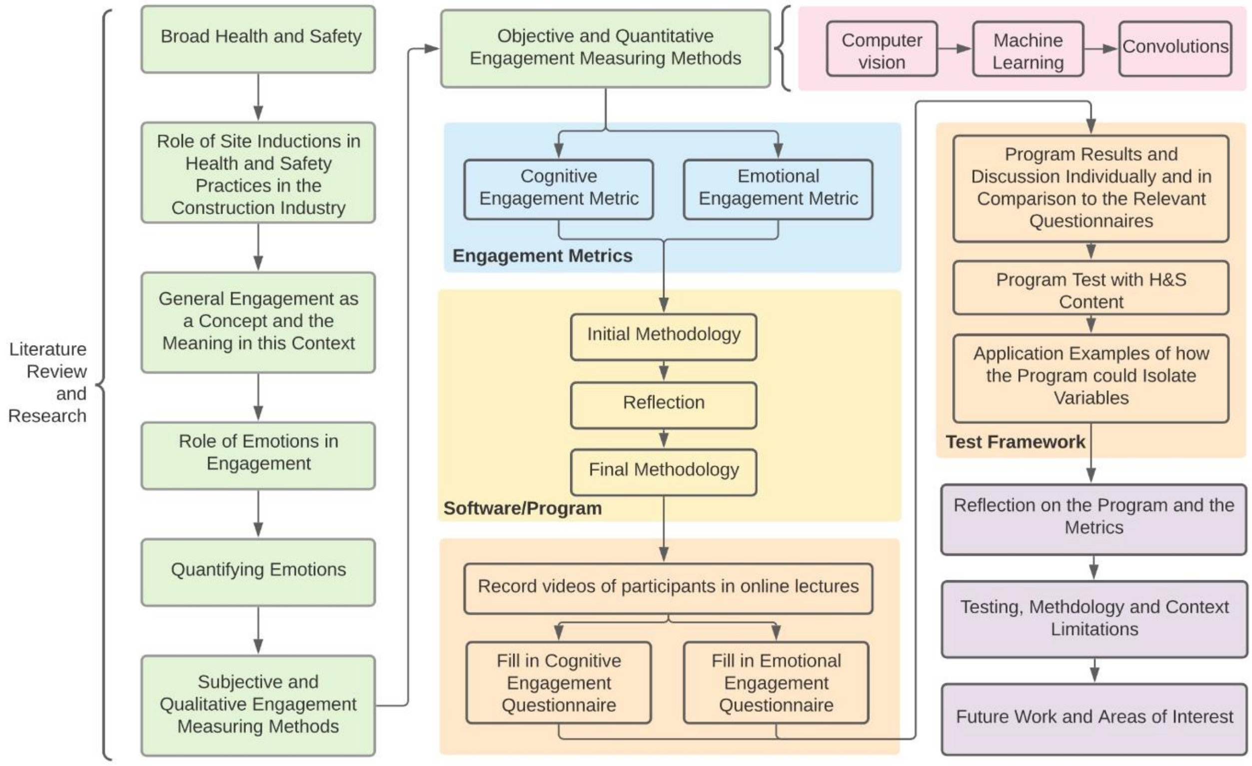

Figure 1 A flowchart showing the general approach to this project

2.0 Background and Literature Review

2.1 Broader Context of H&S

A construction site is widely known to be one of the most dangerous places to work; rates of fatal injuries in this field are approximately four times higher than average rates across all main UK industries, spanning from agriculture to transportation to manufacturing (Health and Safety Executive, 2019).

Physical risks have always been present on site, but a new awareness of the various mental health issues highlights that there is an increasing amount of psychosocial risks faced by the construction work force. Studies and existing literature suggest that there is a reduction in work efficiency as well as a higher risk of accidents and injuries on site for those suffering from mental illnesses (Beseler and Stallones, 2009; Boschman *et al.*, 2013). Despite this, the general attitude towards H&S procedures is underwhelming and the industry lacks a proper safety infrastructure that motivates their employees to work safely (Langford, Rowlinson and Sawacha, 2000a).

Due to the nature of the work, those employed on construction sites often work independently and intuitively on a variety of potentially dangerous tasks (Baxendale and Jones, 2000). The main kinds of accidents occurring over the past three years are slips, trips or falls, injuries while handling, lifting or carrying and struck by moving objects which, when combined account for 61% of construction accidents over a 7-day duration (Health and Safety Executive, 2020). This makes it apparent that most accidents on site occur during routine and repetitive work; this may indicate failures in control and management which is the responsibility of H&S managers as opposed to the result of reckless behavior of construction employees (Baxendale and Jones, 2000).
To ensure that workers undertake all procedures in a safe manner, and hence limit the risks of injury due to general H&S procedural negligence, it is required that comprehensive and tailored H&S programs are prepared in the construction industry (Rowlinson, 2004). According to the most recent 2015 Construction Design and Management (CDM) regulations, a H&S program must reach certain standards and include actions such as producing a thorough set of site rules, safe traffic and vehicle routes, and a suitable site induction.

2.2 Site Inductions and Safety Culture

Of all the various components that make up a successful on-site H&S plan, the focus of this project is site inductions, with the intention of being able to improve this aspect using the proposed method. It has been shown that H&S site inductions correlate positively to safety on site despite no clear explanation of why this is. It is worth considering too the affect time has on these H&S statistics, and whether and whether accident rates increase as time since the last H&S induction or other related H&S site practice. In general, literature on site inductions and their effectiveness is very limited (Harvey *et al.*, 2020). Hence, this project will attempt to partially fill this gap in the literature.

Site inductions are one of the most important but underutilised safety aspects of a construction project (Spillane and Oyedele, 2013) further highlighting that research in this area is needed. Site inductions are legally required in clause 13(4) of the 2015 CDM regulations and are mandatory for anyone who comes on site. The purpose of site inductions are to “impart and educate new operatives of the safety requirements, and so should include basic site information such as the location of welfare facilities and accident reporting procedures alongside key risks and controls such as permits to work, traffic routes and hearing protection zones” (Sherratt, 2016). Current research suggests low levels of engagement from both trainers and workers as well as poor and inconsistent communication of complicated H&S information and lack of discussion of hazards and risks on site (Pink *et al.*, 2010; Harvey *et al.*, 2020). In turn this is likely to lead to poor perception and attitude towards site inductions, especially for those who change site on a near-daily basis and must partake in many inductions (Harvey *et al.*, 2020).

In the past decade, the term ‘health and safety culture’ has become more widely used in the construction industry. This is defined as “the set of beliefs, norms, attitudes, roles and social and technical practices which are concerned with minimising the exposure of individuals, within and beyond an organisation, to conditions considered dangerous or injurious” (Dester and Blockley, 1995). Working towards a more positive H&S culture has put an emphasis on redirecting the aim of everyday H&S procedures to be about increasing worker engagement and instilling the attitude that it is up to workers to take responsibility for their own safety. This would hopefully encourage employees to choose to work safely of their own volition and begin to move away from the traditional approach to H&S which relied on enforced compliance (Langford, Rowlinson and Sawacha, 2000b; Sherratt, Farrell and Noble, 2012). The overall opinion and attitude of the construction industry towards H&S is poor and needs changing (Almond and Esbester, 2016). It has been predicted that even small improvements in the perception of safety could cause dramatic improvements in worker engagement (Whiteoak and Mohamed, 2016).

2.3 General Engagement

It has been seen through multiple studies that increased worker engagement in the construction industry leads to a reduction in accidents (Phelps *et al.*, 2001; Carder and Ragan, 2003; Meldrum and Cameron, 2009). This further supports the theoretical link between engagement at work and workplace safety outcomes (Whiteoak and Mohamed, 2016). Construction workers who are engaged and motivated in the workplace feel more empowered and emotionally committed to identifying and flagging up hazards and concerns that put colleagues and themselves at risk (Lawani, Hare and Cameron, 2017).

A study in the concept of school engagement by Fredricks, Blumenfeld and Paris (2004) explores the multifaceted nature of engagement, and what this means in the context of an academic environment. They split up and define engagement in three different ways: behavioural, emotional and cognitive engagement.

Characterising engagement into these three types is common amongst literature on engagement, particularly in an academic context. Behavioural engagement covers factors like participation, positive conduct, and lack of disruptive behaviour. Emotional engagement refers to the participant's emotional reaction to the activity and the extent of their positive and negative responses. Cognitive engagement is defined as the participant's level of mental investment in the activity (Fredricks, Blumenfeld and Paris, 2004; Ninaus *et al.*, 2019). By thinking of engagement in these three different 'dimensions', it can be seen how engagement varies in length and intensity. This led to the notion that once formed, engagement can begin to build upon itself (Fredricks, Blumenfeld and Paris., 2004) and lead to a better understanding and positive outlook on a given task.

There are many different descriptions of worker engagement; for this project, engagement was initially defined as an invested and committed state of mind where workers are intrinsically motivated, dedicated and absorbed in the task (Fredricks, Blumenfeld and Paris, 2004; Schaufeli and Bakker, 2004; Lawani, Hare and Cameron, 2017). This definition covers both areas of emotional and cognitive engagement; behavioural engagement was not an aspect that this project could assess with only emotion recognition.

Relating Fredricks, Blumenfeld and Paris' (2004) paper to the engagement of construction workers in site inductions has some limitations. The main differences are in environment, learning content and demographic. Their paper is also written with different intentions and focuses- grades and school dropout rates. This means that it is not entirely appropriate assume all the conclusions of this study can be applied to worker engagement in the construction industry. However, site inductions are an educational environment, teaching participants about the risks and safety procedures that must be followed on site. They also rely upon people being engaged and applying their knowledge to their work. The basis of engagement theory in Fredricks, Blumenfeld and Paris' (2004) paper informs work in later stages of this project.

Emotional engagement is based on the emotional responses of the participants which can be directly and comprehensively measured using FER software (Ninaus *et al.,* 2019). However, cognitive engagement is less intuitive to investigate. Plass and Kaplan (2016) describe relevant factors that contribute to this type of engagement, known as cognitive mechanisms. The three basic cognitive mechanisms are memory, attention and perception. This prompted an exploration into the extent to which memory retention plays a role in engagement. Ninaus et al. (2019) describe a set of memory working tasks they developed to see how humans retain information depending on the way that information is presented. Within their article, they discuss a state of total absorption in an activity called 'flow' and how it can have a positive impact on learning that is driven by "a high intrinsic motivation towards the activity that leads to personal positive experiences such as immersion, enjoyment, fulfilment and skill" (Ninaus *et al.*, 2019). Despite this literature being primarily on gamification and not directly related to the H&S industry, it does illustrate another facet to engagement that ought to be considered when later developing a metric to assess if someone is exhibiting expressions indicating engagement.

2.4 The Role Emotions Play in Engagement

A significant factor in one's engagement is emotional response (Fredricks, Blumenfeld and Paris, 2004). The presence of even mildly positive feelings can significantly influence one's thinking (Isen, 2001). Isen *et al.*, 1978 conducted a study to test their hypothesis that a person in a good mood is more likely to retrieve positive material from their memory and that this influences their consequent decision-making behaviour. The experiment was conducted by asking questions to members of the public at a shopping centre where their judgements of products were assessed after a ‘positive mood-inducing event’. Although the setting of this experiment differs to a construction site induction, the outcome of this study states that their hypothesis was true, and that optimism does play a part in positive decision-making processes. This agreed with previous studies in laboratory settings too (Isen *et al.*, 1978).

It has also been stated that positive emotions are more likely to facilitate internalised motivation (Pekrun, 2006; Plass and Kaplan, 2016) and increased autonomy, which is also suggested occurs when individuals find an activity personally significant (Ganotice, Datu and King, 2016). These notions link back to the topic of H&S culture and that workers are demonstrating emotional engagement when they recognise H&S is of personal importance to them. It implies the presence of positive emotions and hence emotional engagement in the H&S inductions can likely impact growth of a positive H&S culture on site.

Emotion has been said to be the foundation of learning (Zull, 2006) since many cognitive processes that contribute to successful learning, like problem solving and decision making, are dependent upon the emotional state of the participant (Chaouachi *et al.*, 2010). Emotions supportive of engagement seem to be those that provoke interest and curiosity, like feeling happy, calm or surprised (Pekrun and Linnenbrink-Garcia, 2012) though these don’t have to be mutually exclusive nor are these exhaustive. The presence of certain emotions is not the only way to understand someone’s engagement in a task. There are many body language cues such as nodding, certain postures and other physiological factors like electrodermal activity and heartrate (Bontchev and Vassileva, 2016) that could infer the response of participants to an event. Only using FER has its limitations in this sense, many studies use a combination of FER alongside a physiological signal measurements to be able to compare results (Gonzalez-Sanchez *et al.*, 2011; Grafsgaard *et al.*, 2013; Kolakowska *et al.*, 2013).

As previously mentioned, both emotional and cognitive engagement are relevant for construction workers in a health and safety induction context. FER software is able to accurately detect emotions displayed which can be interpreted and analysed to understand the extent of positive and negative affect. To understand the link between cognitive engagement and emotions displayed by the participant, it is necessary to explore the relevant cognitive mechanisms, such as memory and attention. It is understood that neutral events are less memorable than positive or negative events and that strength of the emotions elicited is a component in memory modulation (Nielson and Powless, 2007). Studies on animals and humans indicate that the influence of the amygdala upon the hippocampus is critical for memory consolidation (Fastenrath *et al.*, 2014). This was explored further using functional magnetic resonance imaging (fMRI) which concluded emotional reactions modulates amygdala activation in the hippocampus. The outcomes of this study further support that the presence of emotion in any given event to contribute to its memorability but also indicates that emotions can influence memory independent of their strength. Further neuroscience research conducted by Rutherford and Lindell (2011) concluded that specifically, changes in emotional state influence higher cognition.

Both clinical and experimental observations over the past decades have suggested that memory is not a unitary process, but consists of two different forms; short term which can last up to a few hours and long term which can last for years (Goelet *et al.*, 1986). In the context of H&S inductions, the aim is imprint long-lasting memories. Buchanan (2007) identifies emotions as effective retrieval cues for long-term memory Plass and Kaplan (2016) suggest there is a strong case that these mechanisms are “inherently motivated by virtue of their interconnectedness with emotion”. This indicates a strong link between long term memories and emotions.

In addition to the emotions expressed by a person being indicative of their engagement in an activity, there is also experimental research supporting the notion of “emotional contagion” (Bakker, Albrecht and Leiter, 2011). This term describes the tendency to mimic facial expressions and body language of those around you and “converge emotionally” (Hatfield *et al.*, 1994). A study in leadership effectiveness by Damen (2007) explored the change in quantity and quality of a task done by a group with an engaging and enthusiastic leader and an unengaging leader. The group that had the engaging leader completed a higher quantity of the given task compared to the other group (Damen, 2007). This supports the emotional contagion hypothesis. It also leads to the notion that displays of engagement by a participant may propagate more engagement for accompanying participants as well as themselves.

2.5 Quantifying Emotions

The aim of this study is to develop an interpretation of FER outputs that can be deployed effectively during construction site H&S inductions to pin-point events that induce engagement.

To do so, the selected methods will focus on:

- Easily quantifiable outputs enabling the generation of conclusions with minimal uncertainty.
- Measurements of engagement during sessions allowing the isolation of engaging components of inductions.
- A method that is non-invasive and applicable to all participants, maximizing the amount of data captured.
- Selecting a method with a high level of invariance to setting.
- Easy to use by other students/researchers/professionals in the field.

Measuring human emotions accurately is nearly impossible- the definition of the word “emotion” is a contentious topic (Bagozzi, Gopinath and Nyer, 1999; King and Meiselman, 2010; Kenney and Adhikari, 2016). In this study, the set of characteristics defined in Ekman's (1999) ‘Basic Emotions’ will define “emotion”. This consists of six basic emotions: fear, anger, joy, sadness, disgust, and surprise. Ekman claims that these are the ‘building blocks’ of all other emotions.

Many studies contest the use of this definition, whilst others suggest improvements (Jack, Garrod and Schyns, 2014; Gu *et al.*, 2016; Ortony and Turner, 1990; Zheng *et al.*, 2016; Donovan *et al.*, 2020). However, its widespread use in the fields of computer vision (Minaee and Abdolrashidi, 2019) and in studies of engagement (Teixeira, Wedel and Pieters, 2012; Ninaus *et al.*, 2019) make it an accepted tool in this field. This study prioritises reproducibility, and practicality which are both characteristics of Ekman's method that outweigh its potential limitations.

2.5.1 Subjective and Qualitative Methods

The prevailing method of recording emotions and experiences is through interviews and questionnaires. These provide insight into the underlying complexity of human experiences (Seaman, 1999; Dowling, Lloyd and Suchet-Pearson, 2016; Almeida *et al.*, 2017).
Questionnaires can be tailored to their subjects, and are found to confer context, as opposed to objective measures such as physiological markers (Hamilton and Finley, 2019). Typically, they are composed of a small set of questions which can be completed rapidly, making them a relatively fast way of collecting data. The extensive body of literature to compare results against makes them relevant and guides their interpretation.

Qualitative methods require careful consideration of the many confounding factors, for example participant mood, and the phrasing of the questionnaire (Falchikov and Boud, 1989; Spinelli *et al.*, 2014).
As well as their potential to be handled effectively by expert researchers, they are prone to misuse by less experienced teams (Almeida *et al.*, 2017).
Although certain aspects of human complexity can be encapsulated in qualitative measures, they are also subjected to inter- and intra-human variation (Falchikov and Boud, 1989; Lishner, Cooter and Zald, 2008) meaning drawing definite conclusions is often difficult.

As this work is inspired by (Ninaus *et al.*, 2019), comparable methods will be used throughout for ease of interpretation.
(Ninaus *et al.*, 2019) use the well-established Positive Affect Negative Affect Schedule (PANAS). PANAS discriminates accurately between the different affective states (emotional display) of subjects. It is one of the simplest and most explicit questionnaires (Watson, Lee Anna and Tellegen, 1988), making it suitable for undergraduate use and rapid data collection.

The questionnaire is composed of five questions, which address a participant's perceived present engagement, effort, persistence and how absorbed they are in the task.

Although the authors are not aware of its use in the context of the construction industry, its extensive testing meant that the questionnaire was deemed suitable to be used to compare cognitive engagement results in this project.

In addition to PANAS, a separate questionnaire relating to perceived engagement was required. As such a 4-item cognitive engagement questionnaire (CEQ) was used (Rotgans and Schmidt, 2011).

This questionnaire was developed with the intent of being used in a variety of settings and has undergone rigorous validation, thus suggesting it is reliable and appropriate to the context at hand.

2.5.2 Objective and Quantitative Methods

The need for quantitative methods arises from the limitations set out in the literature with regards to assessing human experience through questionnaires (Leue and Beauducel, 2011) PANAS is accepted as an accurate metric but not a precise one (Thompson, 2007). Moreover, it does not provide the granular detail required to isolate single events triggering engagement during site inductions.

The primary aim of developing this method is to eliminate the noise and refine the interpretation of human emotion, thus enabling the linking of engagement to individual and across sessions. The proposed mechanism for this is computer vision.

Using qualitative and quantitative data in combination provides cross examination for each result type. The use of both in tandem compensates for the limitations of each method, thereby increasing the potential for significant conclusions to be drawn. As non-invasive quantitative methods are relatively new in this field (Ninaus *et al.*, 2019), the seniority of qualitative methods provides a robust quality control through existing benchmarks in the literature (Shouse, 2005; Kenney and Adhikari, 2016).

With this in mind, an FER software based on Ekman's Basic Emotions will form the basis of the quantitative method. It provides second-to-second feedback on the emotions of participants in an objective fashion, without the complications associated with physiological data collection.
This software enables collection of large amounts of data rapidly, as it is easy to operate. This is key in attempting to pinpoint what makes site inductions engaging- more data means stronger evidence and less uncertainty.

Basic Emotions are suitable as they have precedence in the field of FER (Susskind *et al.*, 2007; Goldberg *et al.*, 2019; Hui Ma and Turgay Celik, 2019). This means any results can be compared to the literature.

The main limitation of using FER is based in the assumption that affect is reflective of emotions that are felt. This is somewhat untrue. It is known that being observed reduces the emotions displayed (Noah, Schul and Mayo, 2018). In particular, the act of being recorded or watched is known to effect behaviour (Yu *et al.*, 2015) and may even raise one's self-awareness (Baltazar *et al.*, 2014). As such it should be noted that there is a possibility that participants' facial responses may be skewed.

2.6 Computer vision and its applications to emotion recognition
Computer vision's attractiveness comes from its ability to yield rapid, consistent and objective results. Examples range from detecting diabetes (Gulshan *et al.*, 2016) to the development of self-driving cars (Amini *et al.*, 2019).

Although FER can be traced back to the late 1990s, its practical application correlates with the increasing volume of data available. (Fergus, 2015)
An example is the development of real-time FER (Xu, Zhang and Wang, 2017) and the use of radio waves to detect emotion (Zhao, Adib and Katabi, 2016). In industry, companies such as Microsoft provide business services based on FER (Microsoft Cognitive Services).
The use of these technologies has historically been either economically or technically prohibitive. As demand for its use grows, the development of open

source alternatives is expanding (Arriaga, Plöger and Valdenegro, 2017; Tadas Baltrusaitis *et al.*, 2018). In industry, its applications are often limited by the ethical ramifications of the use of such technology (Martinez-Martin, 2019) and the public perception that it is invasive.

Current research surrounding the use of FER to monitor engagement is minimal. Few studies provide convincing evidence for its practical applications (Ninaus *et al.*, 2019).

Three main ways of classifying facial emotions are prevalent in the literature: Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), and Facial Action Coding System (FACS).

SVMs involve the use of 'standard' machine learning (ML) applied to face tracking software (Abdat, Maaoui and Pruski, 2011). In this method, a statistical model is given data from another software with regards to the position of different elements of people's faces in images (e.g. gaze, head tilt). It is also provided with their corresponding emotions (as determined by humans). Then, classification analysis allows the machine to infer what emotions are present in new images based on both of previous data types it has seen.

The use of convolutional neural nets- or deep learning- is a method in which a model is provided images which are pre-labelled with emotions. The model finds out what features are relevant on its own and develops the tools to identify them (Azcarate *et al.*, 2005). The final step is like SVM in that it returns a probability of belonging to a category.

The oldest and most used method is Ekman's Facial Action Coding System Action Units (FACS AU) (Lien *et al.*, 1998). This method is automated insofar that the computer tracks different parts of the people's faces. Hundreds of different movements of parts of the face (action units) are manually ascribed a certain value or emotion according to the method described by Paul Ekman (1983). The computer then returns the value ascribed when it detects certain combinations of action units. Table 1 Comparison of quantitative methods to measure emotion highlights the main pros and cons that were considered during the selection of an appropriate method.

Table 1 Comparison of quantitative methods to measure emotion

Method Types	Advantages	Disadvantages
Support Vector Machines (SVMs)	- Uses simple statistical analysis (e.g. regression) meaning results are easy to understand (Boser, Guyon and Vapnik, 1992) - Simple to maintain and update with new data. - Computationally cheap. Takes minutes to train. - Does not rely on underlying psychological theory that may one day become outdated. - Provides a numerical value of uncertainty for its predictions, thus providing understanding of its limitations to the user.	- Uses input from face tracking software (Abdat, Maaoui and Pruski, 2011). It will be limited by the quality of data it receives from this. - Based on data labelled by humans which is inherently subjective. - Requires millions of data points to start producing accurate results (Susskind *et al.*, 2007). - Take a long time to predict once trained.

Convolutional Neural Nets (CNNs)	- Based on SVMs, so has theoretical precedence. - Does not require manual/external software labelling of features. It will learn these on its own (Azcarate *et al.*, 2005). - Highly generalisable, meaning it can be used in settings different to that it was trained in (e.g. induction rooms). - Provides a numerical value of uncertainty for its predictions. - Forms the basis of all state-of-the-art in FER (Chollet, 2017). - Once trained, can predict emotion in milliseconds - Can be re-trained for new tasks or improvement.	- Based on data labelled by humans. - Requires an even larger dataset than SVMs to reach higher than human level of accuracy. - Contains orders of magnitude more parameters than other methods, incurring large computational and time costs to train the models. - High computational and time cost imposed by training affords less time for optimisation and iteration.

Facial Action Coding System Action Units (FACS AU)	- Has precedence in the literature. Easy to compare results with. - Simple underlying system. Tracks different parts of people's faces. - Ascribes values to different movements of people's faces and returns them to users.	- Based on Ekman's psychological theory from 1978. Possibly outdated in some respects. - Low level of adaptability. Relies on replicable lab-like conditions so not suitable for use on construction sites. - Costly. Requires training.

Following from this, CNNs were identified as the most suitable option. Their ability to deliver accurate results in a variety of settings, to be retrained if needed and to compute results rapidly once trained set them apart as best suited to this project.

2.7 The Theory Behind Emotion Recognition in Machine Learning

2.7.1 Computer Vision

Similarly to the human vision system, computer vision is based on sensing and processing light (Figure 2). Some theories suggest that human processing of vision relies on pattern recognition to predict what objects in our vision are (DiCarlo, Zoccolan and Rust, 2012). The implementation of computer vision follows a similar structure (Susskind *et al.*, 2007).

In digital form, an image is represented as a matrix of pixels (Figure 3). Each

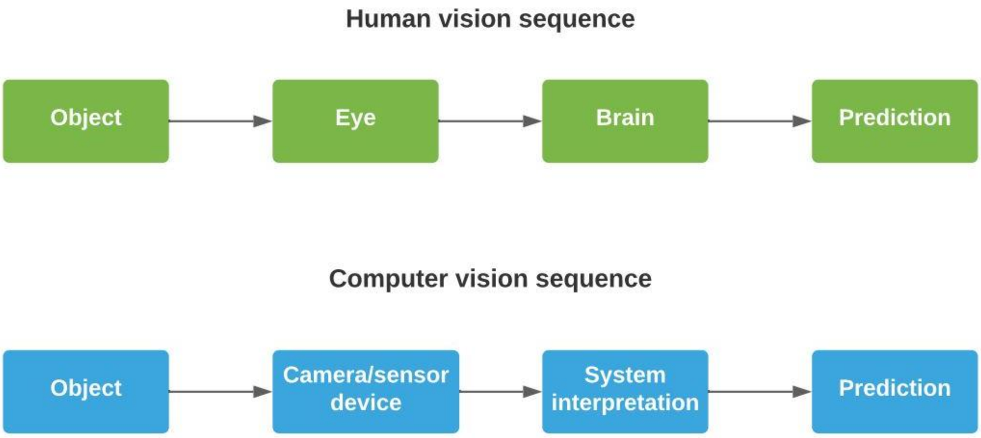

Figure 2 Representation of general human and computer vision sequence (Susskind et al.,2007)

cell of the matrix represents a portion of space in an image. In its most basic form- a grayscale image, the pixels are assigned a value from 0 to 100. The value of each pixel represents the intensity of brightness in it.

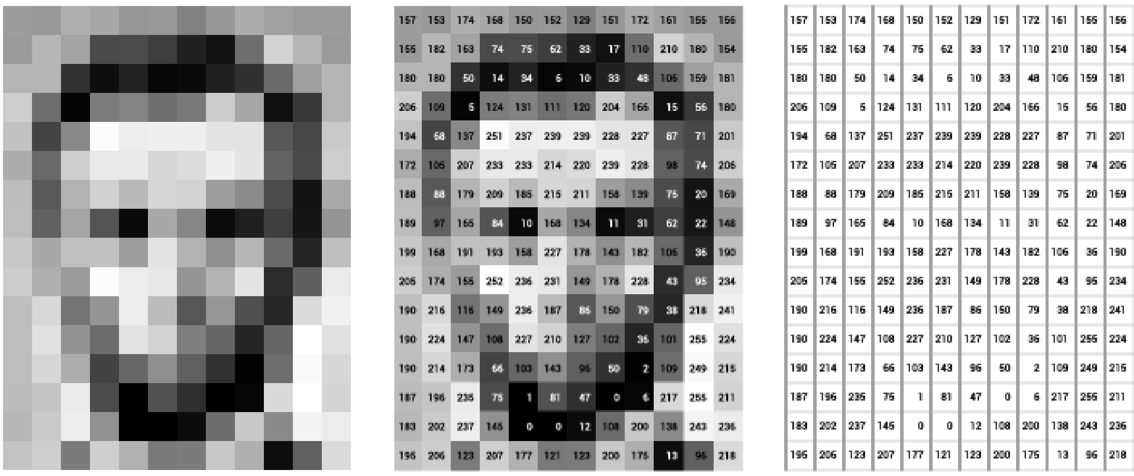

Figure 3 Visual representation of an image to matrix of pixels and numeric values (Levin, 2020)

For coloured images, each pixel has three values assigned to it: red, green, and blue. The intensity ranges from 0 to 255. Any given combination of these can produce a range of colours on a screen. As each pixel is assigned three values, we can define a 3-layered matrix where each layer represents one of the colours.

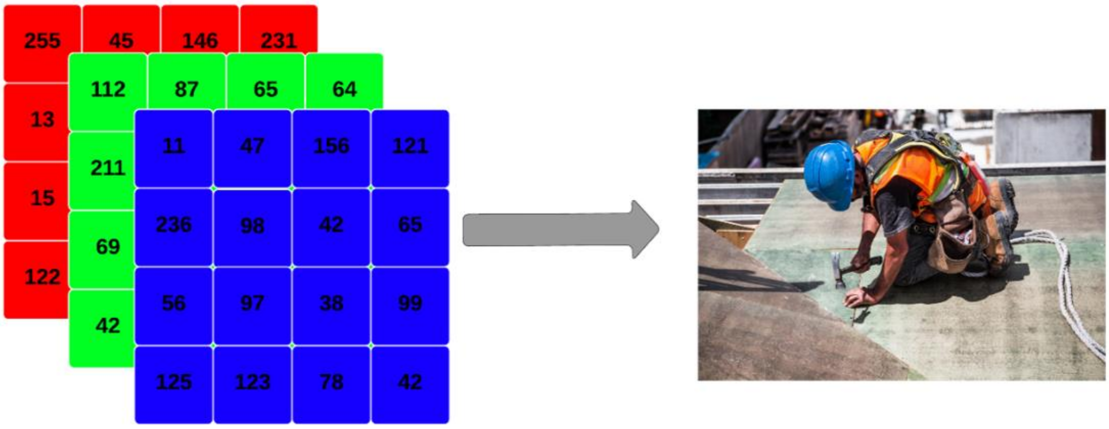

Figure 4 Graphical representation of the three colour matrices that compose a screen image

2.7.2 Convolutions

Computer vision- and more specifically image classification, is based on the manipulation of these matrices using mathematical functions. Using these, features of an image (e.g. a wheel) can be 'extracted' from images. In machine learning, the system by which a series of functions are applied to some inputs and transformed into a single output is referred to as a neuron or a perceptron (Figure 5) These units can be replicated and connected to form a neural net.

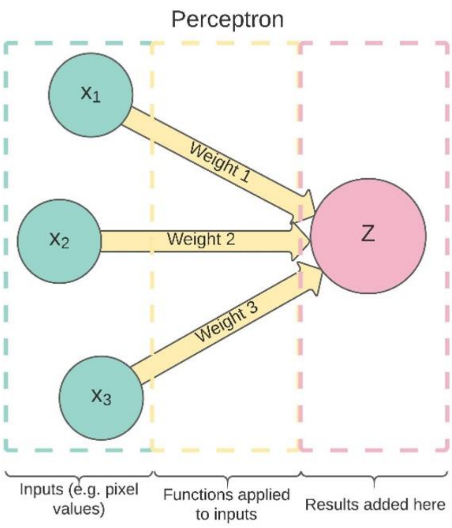

Figure 5 Structure of a perceptron unit

The function of interest in the case of FER is a convolution operator matrix. Their application to imaging is best illustrated through an example.

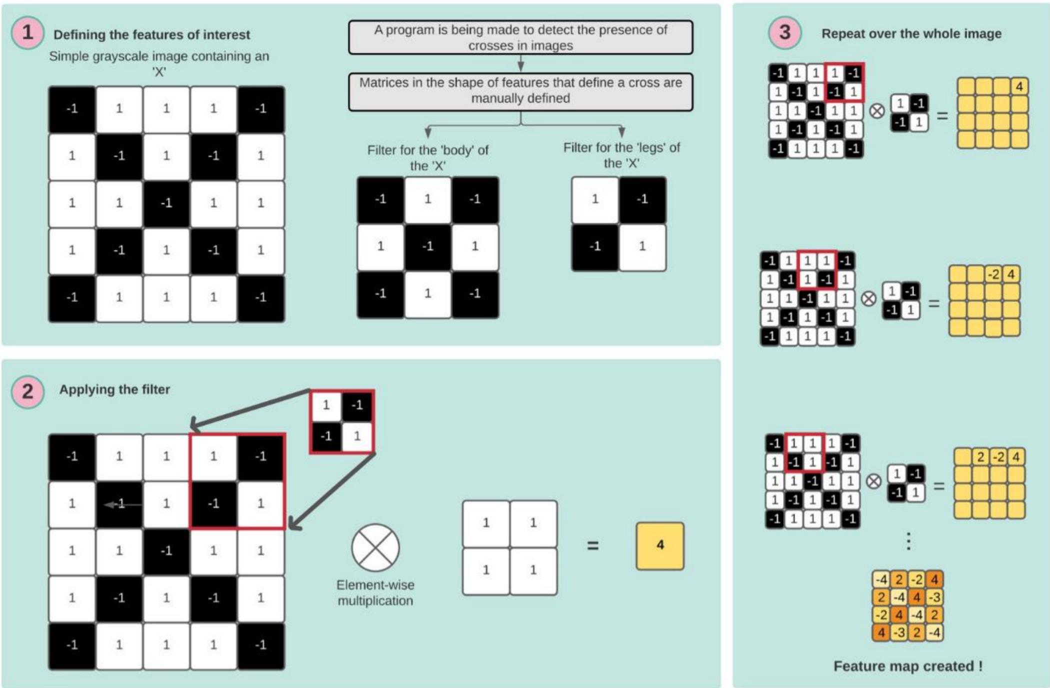

Figure 6 The method by which a feature map is created

With this method (see Figure 6), a new filter can be applied for any given feature. The feature map indicates the strength of the overlap between filter and image, and thus the probability a feature is present in an image. With many features one can begin to infer the presence of objects. In coloured images, the process is repeated with the three-layered matrix.

In the standard application of this method, each filter requires manual definition (such as in Figure 7) and systematic application across all images one is interested in (Felzenszwalb *et al.*, no date; Everingham *et al.*, 2007). In practice, the computational and time constraints this entails makes this method impractical.

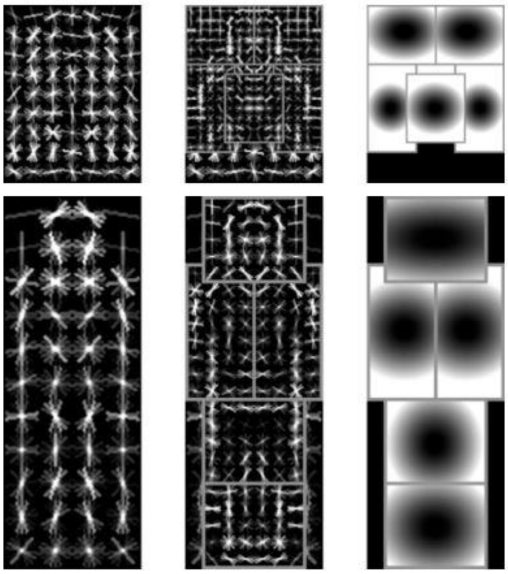

Figure 7. Manually encoded feature extractors for images of a person (Felzenszwalb et al. 2008)

Moreover, it performs poorly on images which contain variation such as changes in viewpoint and illumination (Li, Johnson and Yeung, 2018).
This limited applicability to changing conditions (such as on a construction site) of this method makes it a poor fit for the needs of this research.

2.7.3 CNNs as an Evolution of ML

As an evolution to this method, CNNs use multiple layers of self-defining filters. This is to say, the model starts applying convolutions to the feature maps (see figure 8).

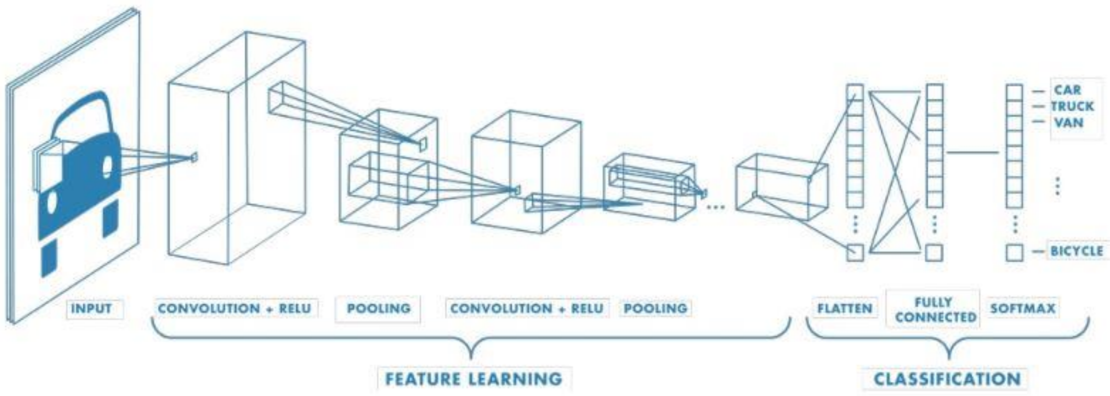

Figure 8 Illustration of how a CNNs multiple layers fit together (MathWorks, 2020)

These self-determining filters are shaped through a process called backpropagation. Using numerical methods and an initial random guess the model produces a marginally better set of filters each time it receives a new item of data. With enough data, many models are able to reach near-perfect accuracy.

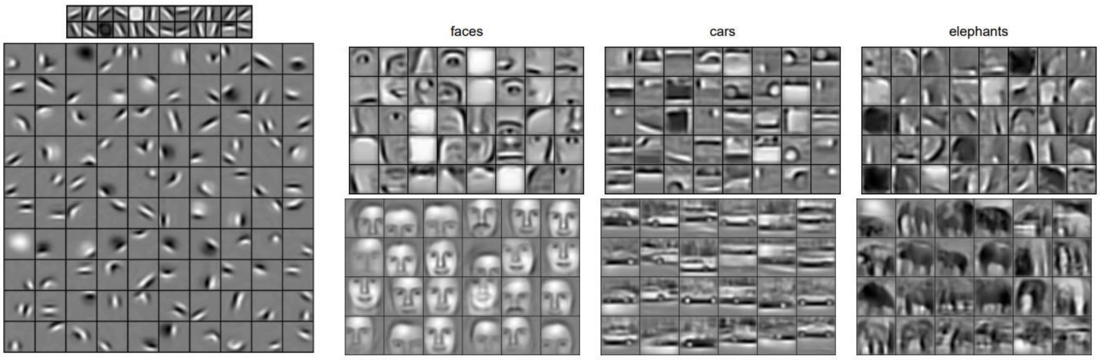

Figure 9 First and second layer of filters of a CNN learning to classify images (left image) and an example of final layer features (three images on right). Taken from Lee et al.,(2015)

It is interesting to note that the filter models get progressively more complex, as can be seen in Figure 9, the initial two filters are usually pairs of lines or circles. In Figure 9, we can see that in later filters, the model is searching for clearly defined features (eyes, noses, wheels, faces)- much as humans might.

Some other additional mathematical operations are required for these models to work (Rectified Linear Unit and max pooling), thus increasing their computational cost in the training phase. CNNs form the basis of all current state-of-the-art methods (Spinelli *et al.*, 2014).

3.0 Work, Testing and Results

3.1 Project Specific Engagement Metric

The literature review informed the approach taken to interpret the data that the FER program alone would output. Following the recording and subsequent application of FER on each session, the result will be a large matrix of percentage strength of each emotion detected per frame. These emotions are fear, anger, joy, sadness, disgust, surprise and neutral. The transformation of this large quantity of data to engagement scores can be seen as a flowchart in Figure 10.

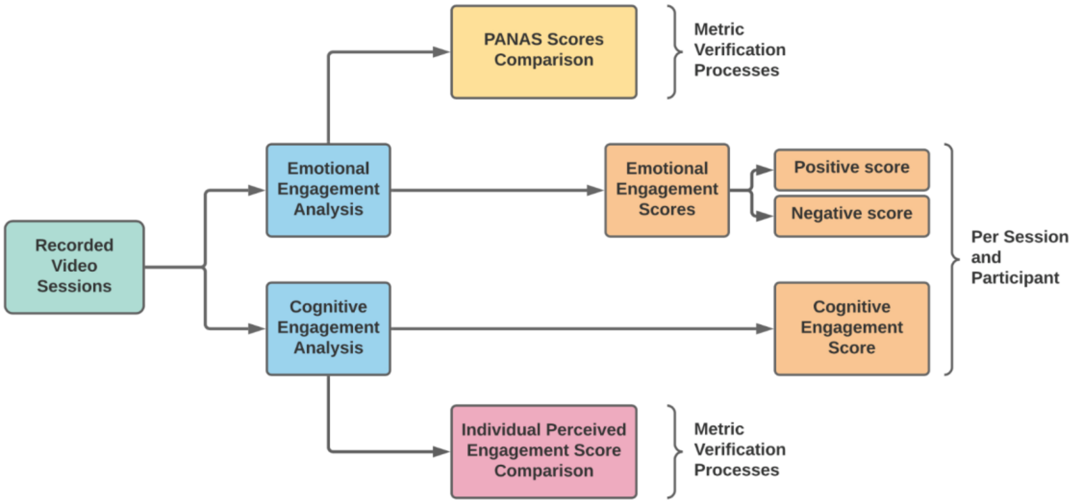

Figure 10 Flowchart showing the process of going from recorded sessions to final engagement scores

Following the research on engagement types and the relevance and correlation of emotions in each, the first step was to begin looking at the presence of emotional, behavioural and cognitive engagement in the context of site inductions. Behavioural engagement could not be measured in this project as it is about participation and lack of disruptive behaviour; the program relies on participants being present and looking at the content or speaker which would infer behavioural engagement by virtue of their involvement. However, emotional and cognitive engagement are based on the presence of positive and negative emotional responses and that of cognitive mechanisms. FER software is able to infer emotional responses which can be further analysed to extrapolate emotional and cognitive engagement.

Despite emotional and cognitive engagement being closely related, the literature review made it clear that they are not directly comparable to each other. This informed our outputs for the program; there are three ‘engagement’ score types for each session and participant. These are based on cognitive and emotional engagement. Emotional engagement is further split into negative and positive scores (Figure 10).

Cognitive engagement will be inferred by understanding the change in emotions expressed by a participant during a session. The literature review indicated that changes in affect generally indicated cognitive engagement as there are correlations between this and the relevant cognitive mechanisms: attention and memory retention. Consequently, this is what will be analysed throughout each session by assessing variance relative to many other sessions. Following each session of the testing phase, a cognitive engagement questionnaire based on the one presented by Rotgans and Schmidt (2011) will be answered. This will act as a point of comparison during the discussion of the results of the testing phase.

Emotional engagement will be based on the relative presence of positive emotions, or lack of negative emotions for a given session and participant. The FER software makes use of 6 basic emotions
(fear, anger, joy, sadness, disgust, and surprise) and of these, fear, anger, disgust and sadness have been labelled negative emotions whilst joy and surprise were positive. Surprise has often been used as a positive emotion in studies, though using it as such involved some level subjectivity (Vanhamme, 2000; Lucas, Diener and Larsen, 2003). This assumption is not confirmed unanimously by literature to be correct (Ortony and Turner, 1990), and is therefore a limitation of this metric.

The program will determine the presence of positive and negative emotions per frame, relative to other sessions. After every recorded session, a PANAS questionnaire will be filled out. This measurement of positive and negative affect will be used for comparison of the test results to see if the results of the metric align with the PANAS results, much like the questionnaire for cognitive engagement. Whilst qualitative emotion measures like questionnaires are subjective and have their limitations as mentioned in section 2.5.1, these two questionnaires have been validated and used in a range of literature. This indicates their results could be useful for cross comparison, not in their absolute values but in the similarity of trends throughout the sessions between these questionnaires and the engagement scores from the FER program.

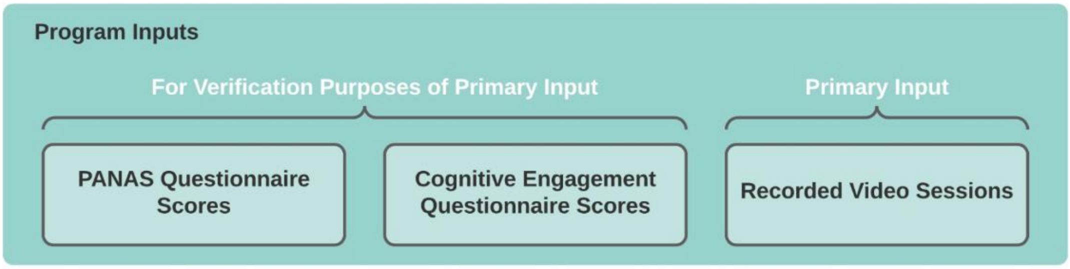

Figure 11 Program Input

The engagement metric requires a video recording of the session as the primary input, and the PANAS and cognitive engagement questionnaires as inputs for the verification of the program (see Figure 11).

The PANAS and cognitive engagement questionnaire scores have been used for verification in this context and do not form an integral part of the cognitive or emotional engagement metric. For someone using the program to assess engagement during a site induction, only the video recording would be required as a program input.

3.2 Methodology

Initially this project aimed to develop its own FER program, deploy it during construction site inductions and collect emotion recordings for participants. Using this data, it was hoped that variables relating to the induction itself (e.g. time of day, type of content) would be reflected in the emotional responses of the participants which could translate into an 'engagement score'. This score would provide a quantitative basis to inform experiments in this area of research, and current induction practices.

As a result of the current pandemic, access to site was prohibited and focus was turned solely to developing a method which could be carried forward in subsequent projects.

Few experiments found in the literature use FER programs to quantify engagement. Of those, almost all are for academic purposes in a classroom setting (Whitehill et al., 2014; Wieckowski and White, 2017; Lasri, Solh and Belkacemi, 2019; Andrejevic and Selwyn, 2020).

3.2.1 Initial Methodology

To meet the objectives set out in section 1, this research initially attempted to build its own FER CNN.

This model was based on the architecture set out in Chollet (2017). It was implemented following the methodology in (Arriaga, Plöger and Valdenegro, 2017)Arriaga, Ploger and Valdenegro (2017).

The main driving factors behind building a model were to retain full ownership of the software, with a view to allow other students and researchers to use it afterwards, and providing the ability to improve the model over time.

Following two iterations, the model was deemed unsatisfactory due to its medium level of accuracy (67%).

For the sake of brevity and relevance, the full model including explanation of the methodology can be found at:

https://github.com/CharlesPlusC/FER/blob/master/27_11_20_CNN_v2.ipynb

```
['Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral']
[[9.9956590e-01 6.9954808e-06 4.2688241e-04 8.5026219e-10 9.2643333e-08
  5.7915446e-09 1.0075586e-09]]
```

Figure 12 Example output of an 'angry' emotion

{'angry': 0.01812676346162334, 'disgust': 2.6663984797892226e-06, 'fear': 0.4790141712874174, 'happy': 79.06026244163513, 'sad': 0.953293
2192087173, 'surprise': 0.012093986151739955, 'neutral': 19.477203488349915}

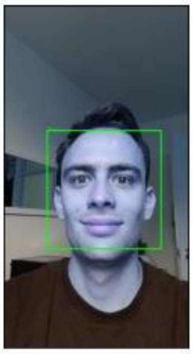

Figure 13 Example output of a 'happy' emotion

Figure 12 shows an example output of the model. As well as returning the emotion ‘composition’ in the image, the model identifies the face and highlights it with a green square. The numerical output is a one-dimensional array of the probability of presence of each emotion emotions in percentages.

3.2.2 Background on DeepFace

Another round of research within the literature revealed that certain models yielded accuracies in the 90%-100% range. OpenFace 2.0 (Baltrusaitis et al.,2018) and DeepFace (Taigman *et al.*, 2014) were of interest as these are open source, making them suitable for the purposes of this research.
These are based on the merging of state-of-the-art models and complex data manipulations to generate predictions. Moreover, they were trained on much larger datasets (tens of millions of images) than the ones used in the initial model as a result of their affiliation with large technology companies (e.g. Facebook AI Research). As such it was decided to focus on using these models in practice, rather than building them.

DeepFace was identified as a suitable option. In terms of practical applications, the model was considered on the basis of its applicability to construction site induction settings.

Construction sites vary in physical condition (e.g. illumination, room layout). DeepFace asserts that its accuracy is invariant to pose, illumination, expression, and image quality (Li, Johnson and Yeung, 2018; (Serengil and Ozpinar, 2020)). Although no numerical value is associated with this claim, the model’s overall accuracy of 98.87% (Taigman *et al.*, 2014) was taken as sufficient proof of its effectiveness in different settings.

With the application of this model on site in mind, the potential for high quantities of data to be collected was deemed important; a large sample size would result in statistically significant conclusions.
DeepFace is able to identify and process many faces per frame. This is important as it reduces the number of cameras required (making the experiment more economically viable) and the potential paranoia their presence may induce. Depending on the settings, DeepFace is able to process faces that are at a minimum of 150 x 150 pixels in size. Considering a standard ‘high resolution’ image is 1080 x 1920 pixels, this means around 84 faces could theoretically be processed in an image, and so a single camera would likely be sufficient in for use in an induction room.

Figure 14. Application of DeepFace to an image with multiple faces present- as might be the case in a site induction. The percentage score displayed above the boxes represent the model's certainty that the faces are faces. (Image from Guardian News,2018)

3.2.3 Data Collection and Exploratory Data Analysis

3.2.3.1 From Video to CNN Output

As a result of COVID-19 restrictions, data of site induction participants could not be collected. The nearest available source of similar data to that of participants in a site induction were recordings of the authors during various online lectures.

A set of data comprised of 27 recordings was gathered to form the basis of the analysis that would test the hypothesis that engagement could be quantified using facial emotion recordings.

For 2 months, all online lectures that were attended were recorded. In addition, a PANAS questionnaire and a cognitive engagement questionnaire were filled out immediately following the lectures. Throughout the lecture, timestamps of different ‘events’ occurring were noted.

These included events such as change in material presented (e.g. video to ppt), beginning of discussions, technical problems, etc. Finally, the type of content and any defining characteristics of a session were noted: Is the lecture pre-recorded? Is this an interactive session? Is the group size particularly small?

Once recorded, the videos were split into their component frames using the OpenCV 4.5.1 library in Python.

One frame was extracted from the video every 30 frames. This equates to one frame every 0.5 seconds (as the cameras used record 60 frames per second). This time interval was selected for two reasons. Firstly, the longest processed videos were around two hours which amounted to around 14,000 frames when processing 1 frame every 0.5 seconds. This quantity of data was on par with the available hardware’s processing limit; a higher quantity of frames proved too computationally expensive.
Through trial and error, it was found that 2 frames per second retained most of the relevant information within the data, when compared to increased frame rates (average emotion, variance). Although emotions that are displayed between frames may be missed, these are nearing the realm of micro expressions which were beyond the scope of this project (Ekman and Friesen, 1969; Ekman, 1991, 2004).

Comparatively, higher frame rates generated little benefit in terms of analysis and significantly slowed processing times. As such, 2 frames per second was deemed a suitable trade-off between processing capability and data quality.

One possible solution to this problem would be to use cloud computing such as Amazon Web Services to gain access to higher computing power.

The frames were then labelled sequentially (e.g. Video1_Frame30; Video1_Frame60; Video1_Frame90; etc…) and stored in a Numpy array, which was passed into DeepFace where an 'emotion vector' was returned for each image. An emotion vector provides values for the presence of each of the emotions in an image (each amounting to 100%). All the emotion vectors corresponding to a video were stored into a Pandas Data Frame for later use.

Table 2 Example 'raw' output from the CNN for the first 5 frames of one session

instance	surprise	angry	happy	disgust	fear	sad	neutral
1	3.604136e-03	0.047555	0.031411	2.053004e-08	0.046440	0.158941	99.712050
2	1.550240e-09	0.049768	0.000002	2.445339e-13	10.007245	89.713948	0.229037
3	1.762349e-04	0.054162	0.006334	3.505967e-07	7.919161	82.924591	9.095575
4	5.667206e-03	0.068729	0.718138	2.871213e-04	34.262314	43.516257	21.428606
5	4.481160e+00	3.324270	0.309641	4.001190e-03	61.850976	26.658156	3.371793

3.2.3.2 From CNN Output to Practical Results

When plotted 'raw', without any modification, the data was illegible (Figure 6).

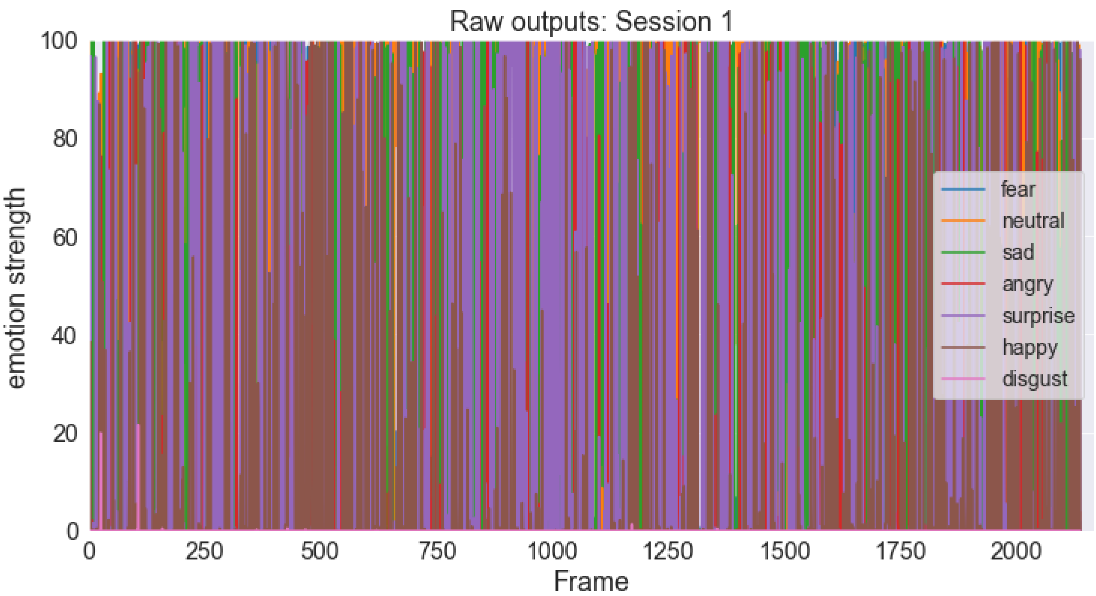

Figure 15 Raw data plotted for video 1 of participant 1. 2 frames were processed for every second of video.

To allow for comparison with existing literature, the emotions were split into positive and negative categories (Ninaus et al.,2019). Reducing the number of variables in each session also made graphical interpretation easier.
The negative category included ‘fear’, ‘disgust’, ‘angry’, ‘sad’. The positive category included ‘happy’ and ‘surprise’. Neutral was kept as its own category. Once split into positive and negative, the emotions in each category were averaged. Figure 16 shows that the data contains many spikes. This is as a result of the discrete way in which the images were processed.
It was deemed that gaining a more ‘continuous’ understanding of the emotions through time was more reflective of human emotion and would be more useful when comparing the data to videos (which are continuous).

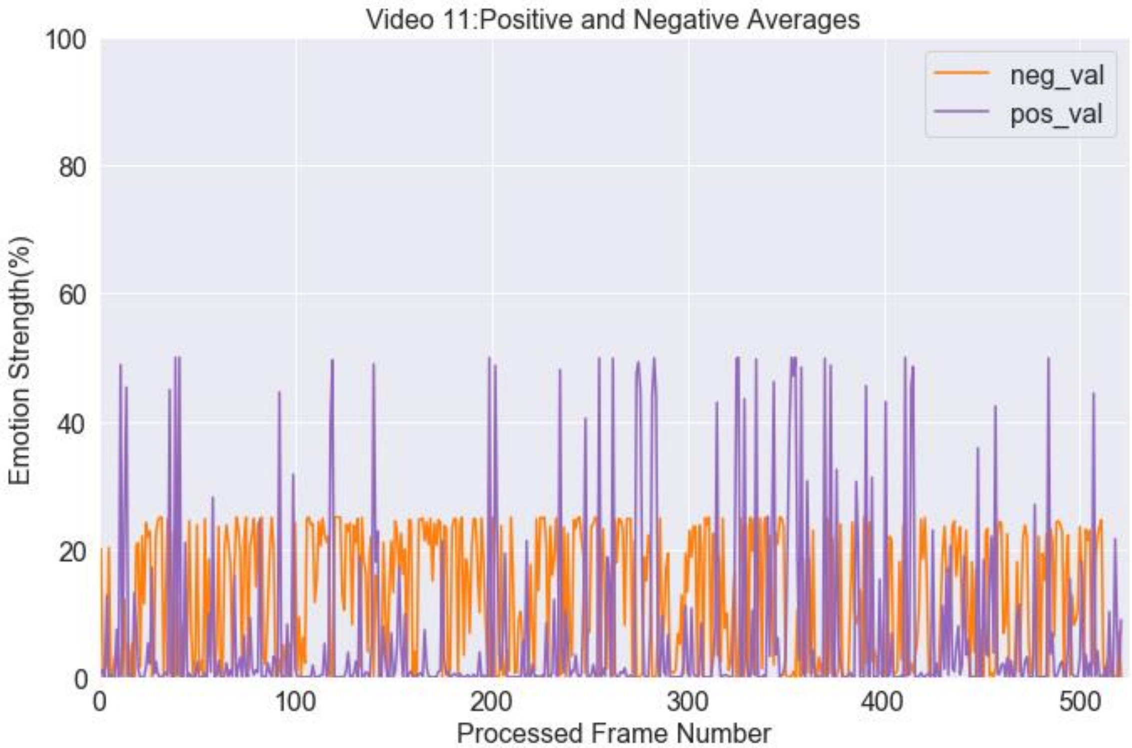

Figure 16. Emotion data for video 1 of participant 1. Positive and Negative emotions have been grouped. Note that the total percentage does not add up to 100% as a significant proportion of the recorded emotion is ‘neutral’.

In order to smooth the data, a rolling average (arbitrary length of 2% which is the length of the overall number of datapoints in any video) was applied to the data. This was successful in providing a clear graph from which quick graphical analysis could be undertaken. While Figure 15 and Figure 16 present the same information, the rolling average is much clearer.

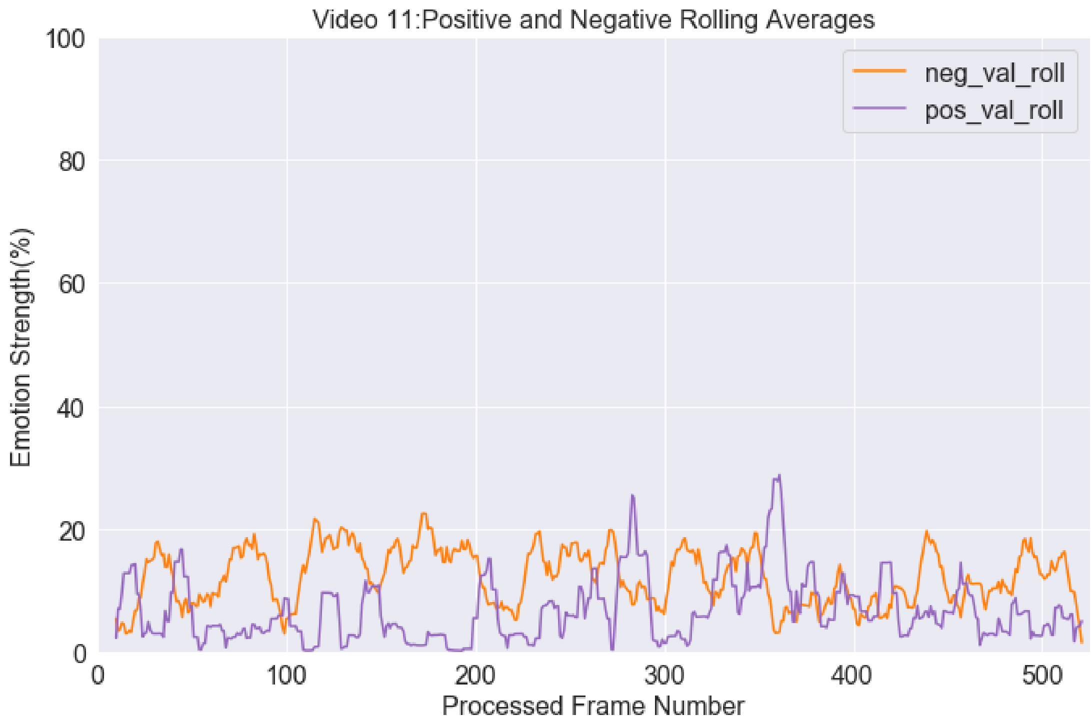

Figure 17. Rolling average of positive and negative emotions for session 11 of participant 2

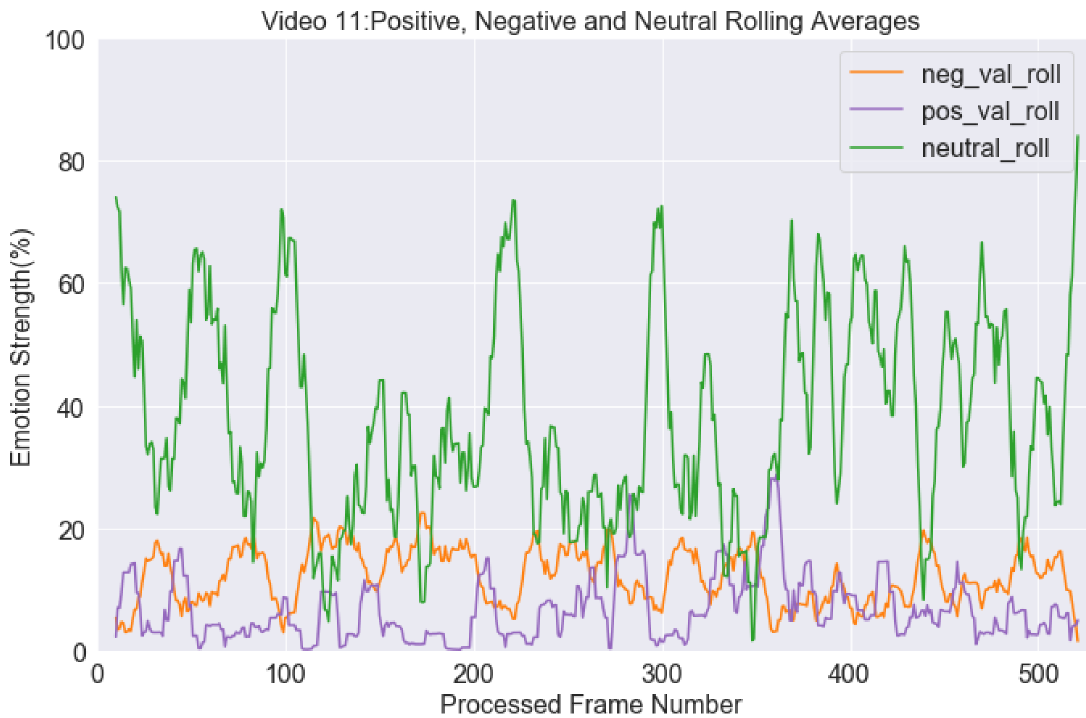

Figure 18. This graph presents the same information as the one above (Figure 17), with the addition of neutral. Note that neutral usually accounts for most of the emotion present during a session

One drawback of applying a rolling average is that emotions that were short in time would be not be as visible in graphs. Overall, this was not seen as detrimental to the quality of the experiment as the rolling average was only used for graphical interpretation.

Figure 19 illustrates the process of going from CNN output to a session graph.

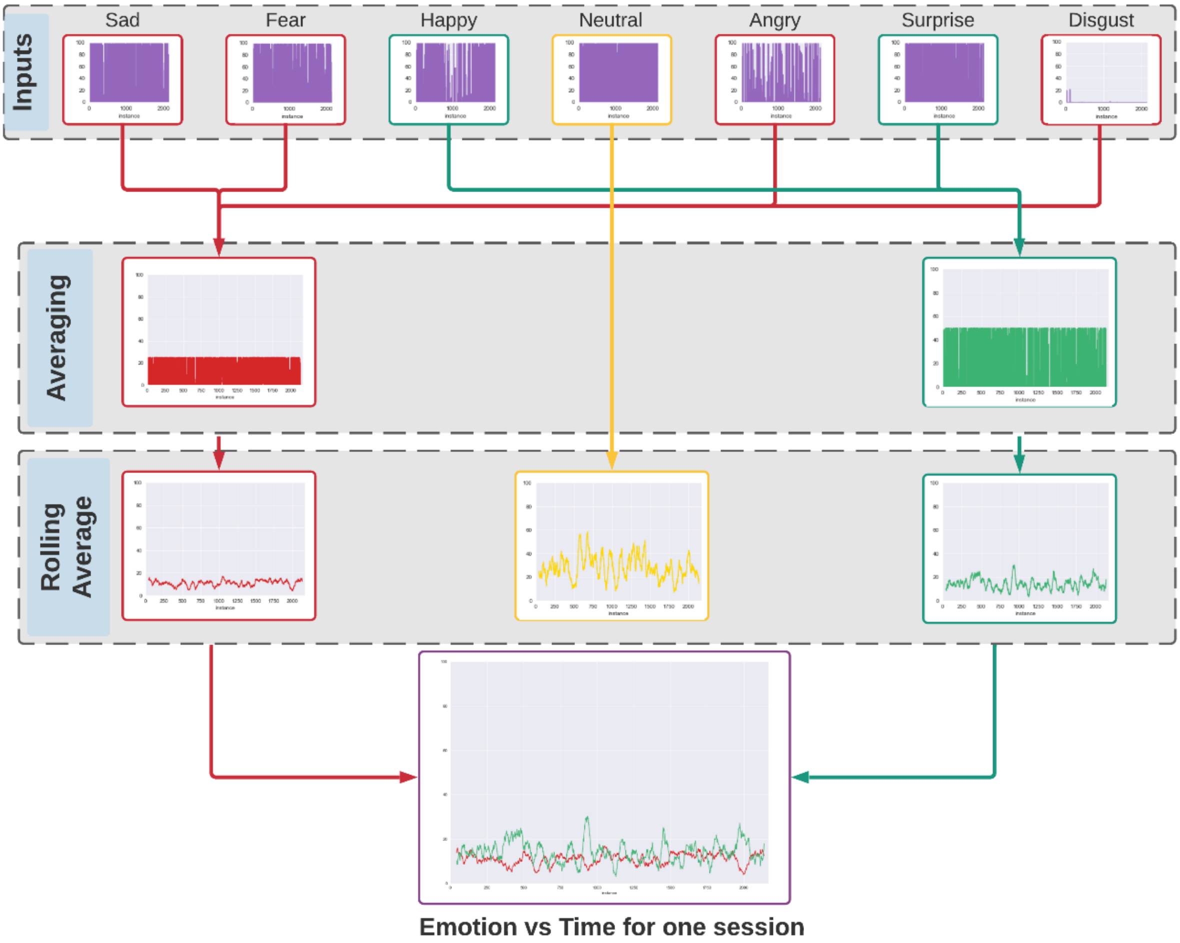

Figure 19. Illustration of how data was taken from 'raw' CNN output to a format in which it could be analysed to identify engaging events during the recorded sessions

3.2.3.3 Exploratory Data Analysis and Quality Check

As a high-level verification of the quality of the outputs, videos were passed through DeepFace and the output compared to the live video. The first video to be checked was a short clip where faces were pulled. The reading was as expected. Figure 20 illustrates the output of this first test and that it was successful in recognizing the correct emotions.

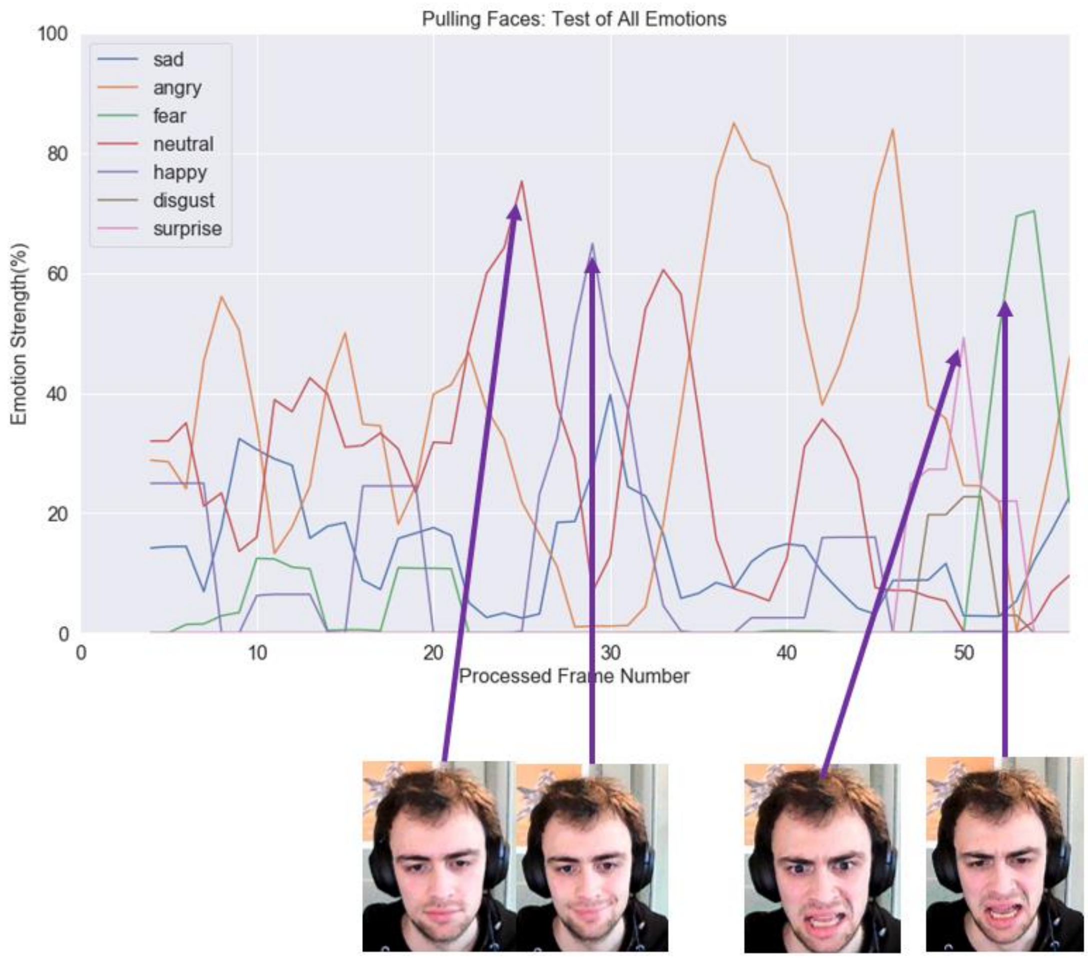

Figure 20 Emotion strength during a short test video. Different faces were pulled with the intention of triggering different outputs

More graph-to-video comparisons were undertaken with a sample of the data from lectures (Figure 21 and Figure 22). Again, the outputs of the CNN aligned with the video.

Although this verification method is somewhat subjective, the effectiveness of DeepFace has already been well established; this verification is only a check on its correct implementation in this context.

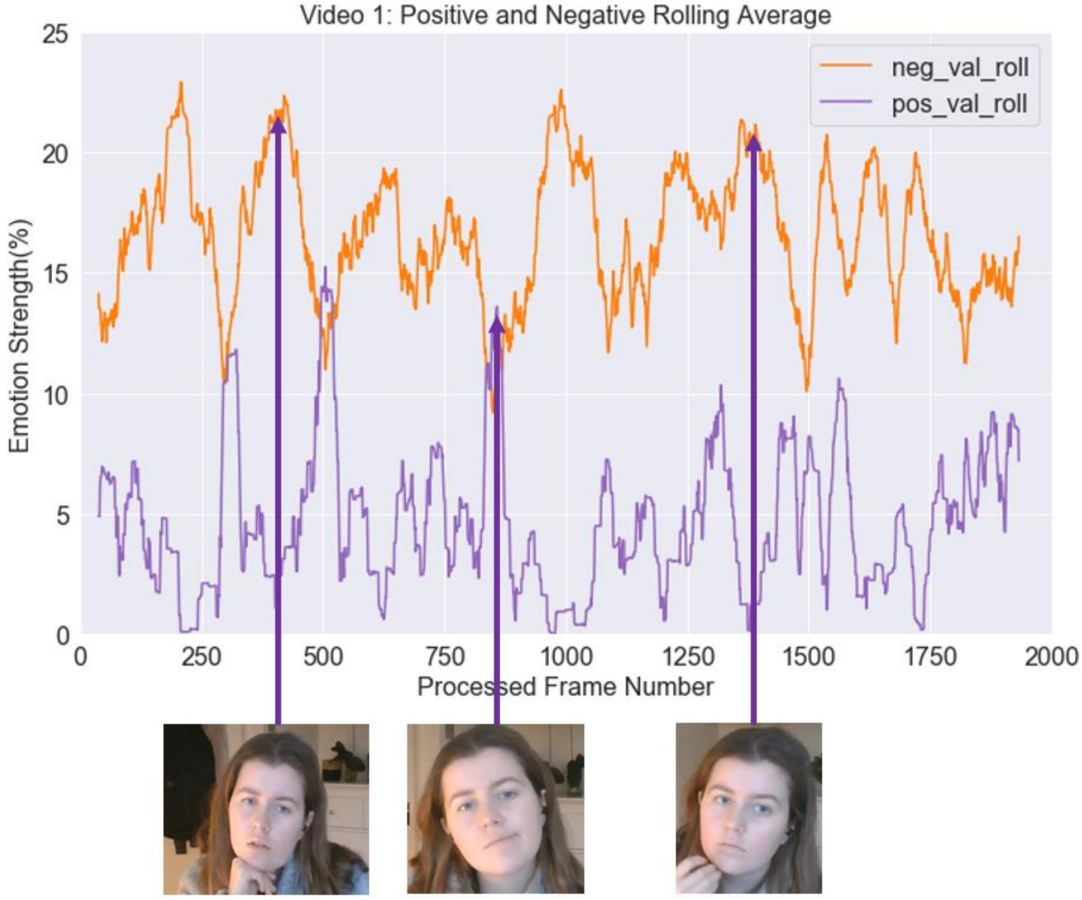

Figure 21 Emotion strength for Video 1 of participant 2:Small Group; Q and A session; Some interaction with lecturer

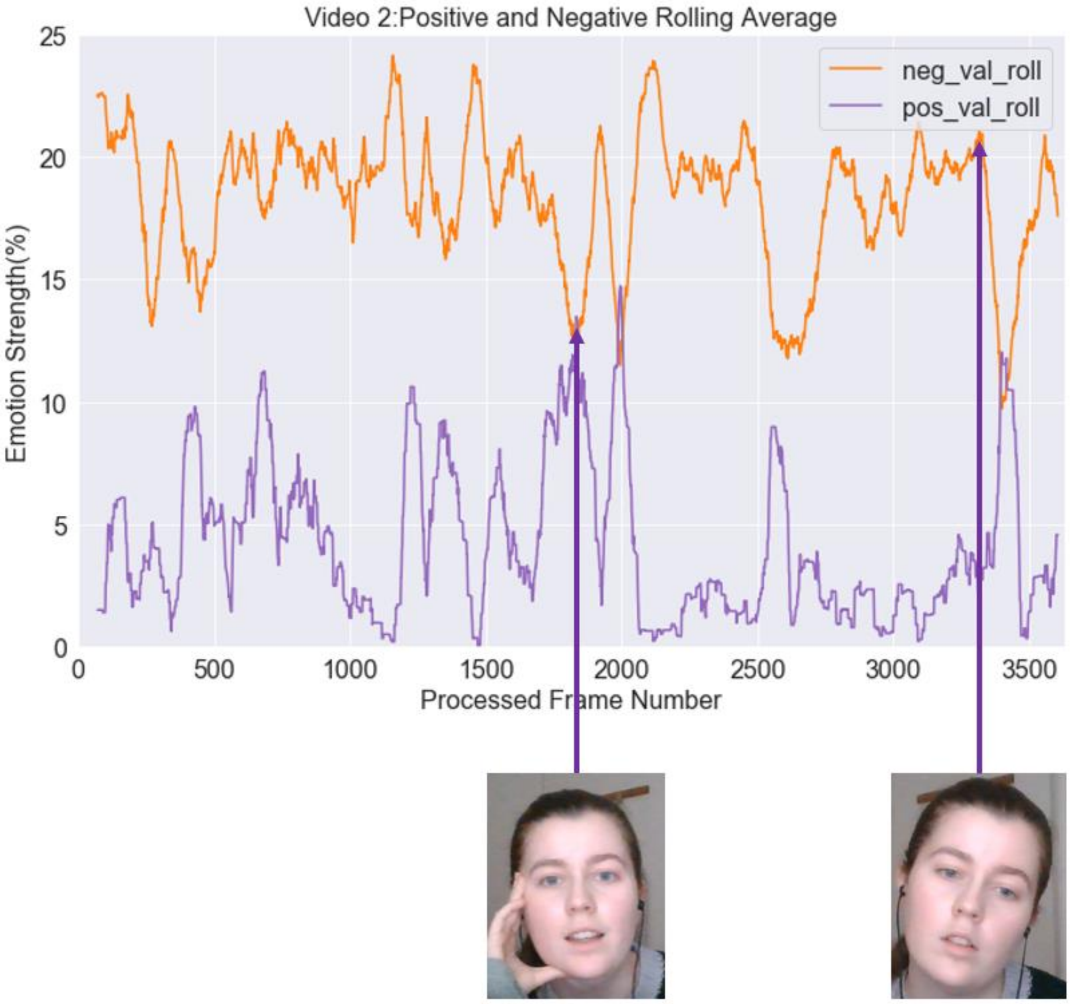

Figure 22 Emotion strength for Video 2 of participant 2: Small Group; One interaction with lecturer around the midpoint of the lecture

3.3 Measures of Engagement

3.3.1 Engagement Data

The aim of this project is to quantify engagement in the context of construction site inductions. It was decided that for the benefit of usability by the layman, the result of this method should convey whether participants are engaged in a straightforward manner, involving minimal analysis.

As such, for any session, a set of three ‘key’ values are provided per participant: difference in variance, difference in positive emotion, and difference in negative emotion. At a glance, these should answer the question of whether a session was successful or not. In addition, further analysis of the data can be undertaken to pinpoint exactly what events triggered engagement and what variables would contribute to a more engaging induction (see section 3.3.5 and 3.3.6).

Although initially it was hoped that individual sessions could be whittled down to a single value, it was found that this was not possible without compromising on the significance of result.

3.3.2 Quantifying Cognitive Engagement

As outlined in section 2.5, cognitive engagement is reflected in the spread of a participants' emotion through time.

In order to quantify the amount of variation a given session contained, variance was calculated for each emotion (except neutral).

Variance is defined as the sum of the squared difference between an emotion at each instance and the average emotion strength for that session, divided by the number of frames minus 1 (eq.(1)). Using variance provides a quantitative grasp on how 'stable' a participant's emotions were during a session.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \tag{1}$$

For each session, the mean of the variances of each emotion was taken. Doing so was at the cost of some of the detail in the data (i.e. exactly which emotions show high variance). However, narrowing down to a single value per session allowed for quick comparison between sessions and identification of which sessions triggered cognitive engagement.
This approach yields a simpler result which still retains meaning.
In fact, due to the squared term in the calculation of variance, even if only one emotion showed high variance compared to the others, it would still be reflected in the 'session variance' score due to the exponentially high score it would yield.

Between both participants, the average variance score for all sessions differed by a factor of almost 2. Evidence exists to explain differences in emotiveness (Doherty, 1997), but none were found which would allow to account for this variable in a numerical way. As such, results were normalized to allow for more objective comparison.

The percentage difference between a participants' average variance (over all sessions) and the variance of each individual sessions was found. In this way, a difference in variance could be found for each video.
This quantified a participants' variance relative to themselves and allowed to determine if a given session contained particularly high amounts of emotional change relative to a participants' 'baseline'.

3.3.3 Quantifying Emotional Engagement Data

Emotional engagement has been defined as a participants' relative average emotion during a session.

To reduce the number of values representing emotional engagement, positive and negative emotions were averaged for each session, giving two numerical values to represent a participants' emotional engagement.

Similarly to cognitive engagement, average positive and negative emotion between participants differed to the point of noncomparability. As such the same process of normalization was applied, and a percentage difference between average positive and negative emotion over all sessions and average positive and negative emotion during each session was found.

3.3.4 Numerical Contextualisation

To further contextualize these two 'metrics' and to get an understanding of their effectiveness, results were compared to independent methods of rating emotional and cognitive engagement.

A Spearman's Rank and t-test (Table 3) were done between the results of the CNN, the PANAS (Watson, Lee Anna and Tellegen, 1988) and the Cognitive engagement questionnaire (Rotgans and Schmidt, 2011). The Spearman's Rank correlation test is a statistical procedure that measures the relationship between two variables (Corder and Foreman, 2011). The value of Spearman's correlation coefficient lies between -1 and 1, where -1 is a very strong negative correlation, 1 a very strong positive correlation and 0 indicating no correlation at all. See eq. (2) for the formula to calculate the Spearman's rank coefficient.

$$\rho = r_S = 1 - \frac{6\Sigma D_i^2}{n(n^2-1)} \text{ (Corder and Foreman, 2011)} \qquad (2)$$

A t-test is a method of performing statistical inference for a set of two groups of data (Seltman, 2018). The t-test analyses the difference between sample means, divided by the estimated standard error of that difference, see eq. (3). This is useful as it infers the probability that the correlation determined by the spearman's rank is correct. The t-test output is a p-value which can range from 0 to 1. A p-value of 0 means the probability of that the null hypothesis is correct is very low, whilst a p-value of 1 indicates absolute certainty.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)}} \text{ (Seltman, 2018)} \qquad (3)$$

To obtain p-values, t must be first calculated and then p-values can be found using statistical p-value tables. The null hypothesis (H_0) and alternate hypothesis (H_1) are as follows:

H_0: Both sets of data are uncorrelated

H_1: Both sets of data are correlated

Table 3 Spearman's correlation rank and t-test score table comparing the outputs of the CNN with independent markers of engagement

Spearman's correlation rank and t-test scores		
Compared Values	Participant 1	Participant 2
Positive Affect and Difference in Positive Emotion	r value = 0.572 p value = 0.084 H_0: Rejected	r value = 0.614 p value = 0.009 H_0: Accepted
Negative Affect and Difference in Negative Emotion	r value = 0.13 p value = 0.719 H_0: Rejected	r value = -0.062 p value = 0.814 H_0: Rejected
Perceived Engagement Questionnaire and Difference in Variance	r value = 0.711 p value = 0.021 H_0: Accepted	r value = 0.779 p value = 0.000 H_0: Accepted

The relationship between positive affect score and average positive emotion for participant 1 and participant 2 indicates a medium strength correlation. However, only in the case of participant 2 was the null hypothesis rejected (although participant 1 is close). It is believed this might be due to the small number of points available for participant 1.

The relationship between negative affect score and average negative emotion is largely uncorrelated. It does not serve as a good verification of the outputs of the program. It is believed this is since negative affect scores were all very low throughout the sessions (Figure 23).

The relationship between the perceived engagement questionnaire and the difference in variance is the strongest relationship for both participants, and in both cases the null hypothesis is rejected.

Overall, the level of agreement between both data outputs was deemed satisfactory for the purposes of verifying that the outputs of the CNN were as expected and comparable to the literature. Some level of discrepancy is to be expected considering the questionnaires evaluate perceived affect and engagement, whilst the program measured displayed affect and engagement.

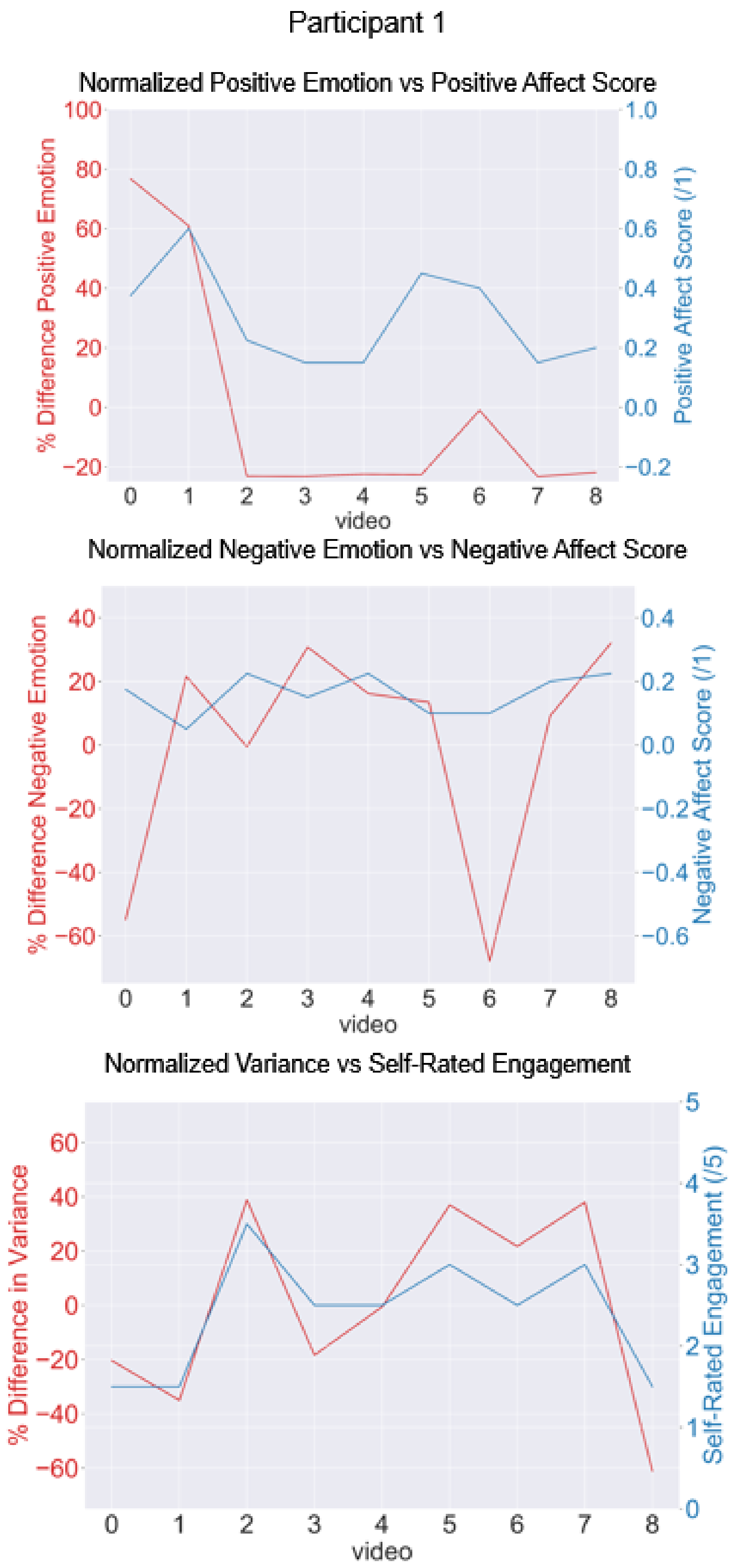

Figure 23 Graphical representation of the program outputs against the relevant qualitative outputs for participant 1 throughout 8 sessions

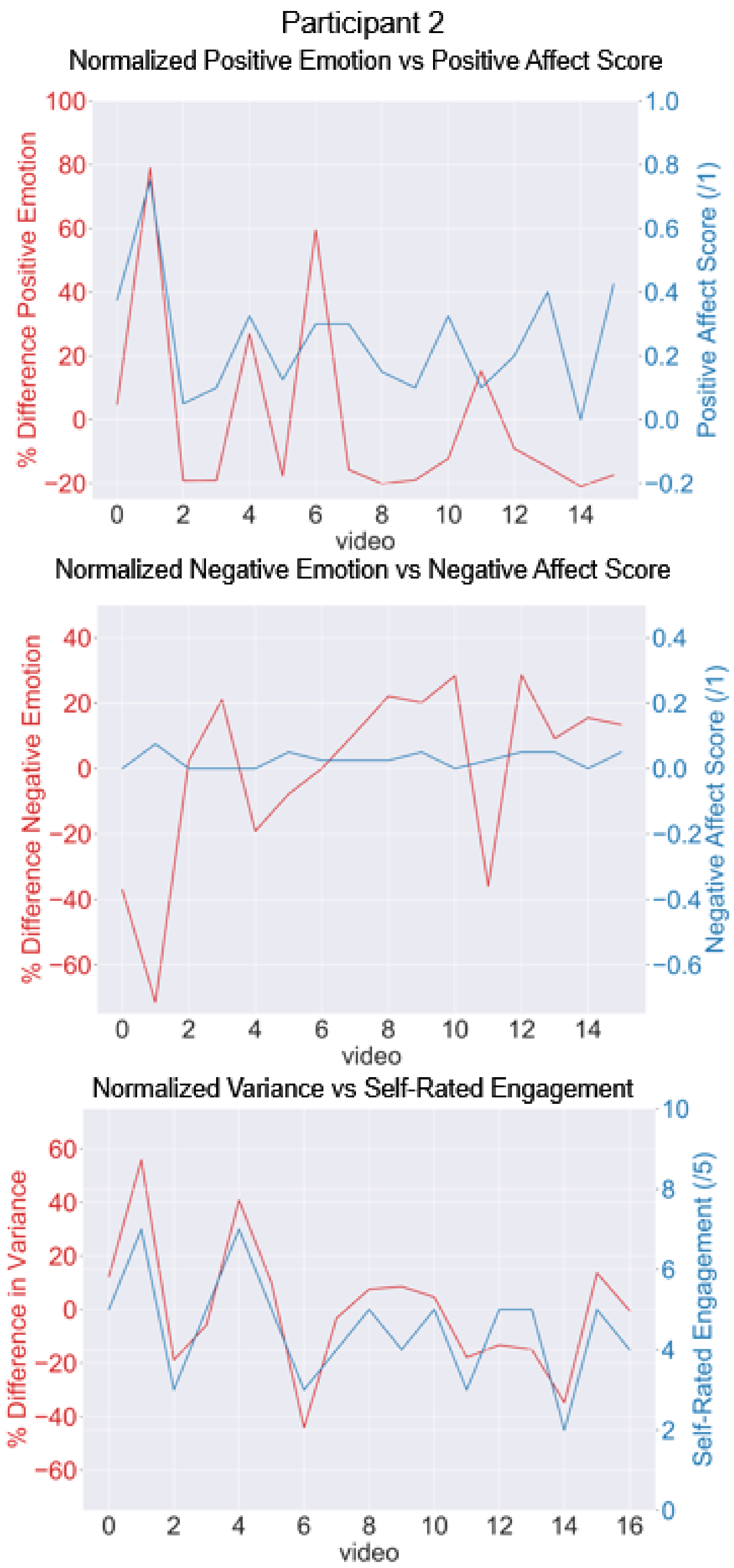

Figure 24 Graphical representation of the program outputs against the relevant qualitative outputs for participant 2 throughout 8 sessions

3.3.5 Application example 1: Isolating engaging events

As stated in section 2.5.2, one of the main aims of this project is to provide a program that may be used to isolate engaging events in the context of health and safety inductions.

In order to address this specific aspect, the authors watched a pre-recorded site induction video from a UK construction site (WJEC Construction, 2016).
Figure 25 shows the outputs of both participants during the video, and their engagement scores.

Both positive and negative emotion strength of both participants follow each other closely during the video indicating the participants responded in a similar manner.
By going over the video with the outputs at hand, it was possible to identify specific events that triggered certain reactions.
This is a useful output as it is believed that with when applied on site, this kind of information would allow the identification of exactly what ‘events’ make a site induction more engaging.

Although this is only a single session, a similar agreement between participants’ emotions was observed during lectures that the participants watched independently. This suggests similar trends in emotional response between individuals can be expected.

This was not investigated further as a sample of size of two would not have yielded any significant results.

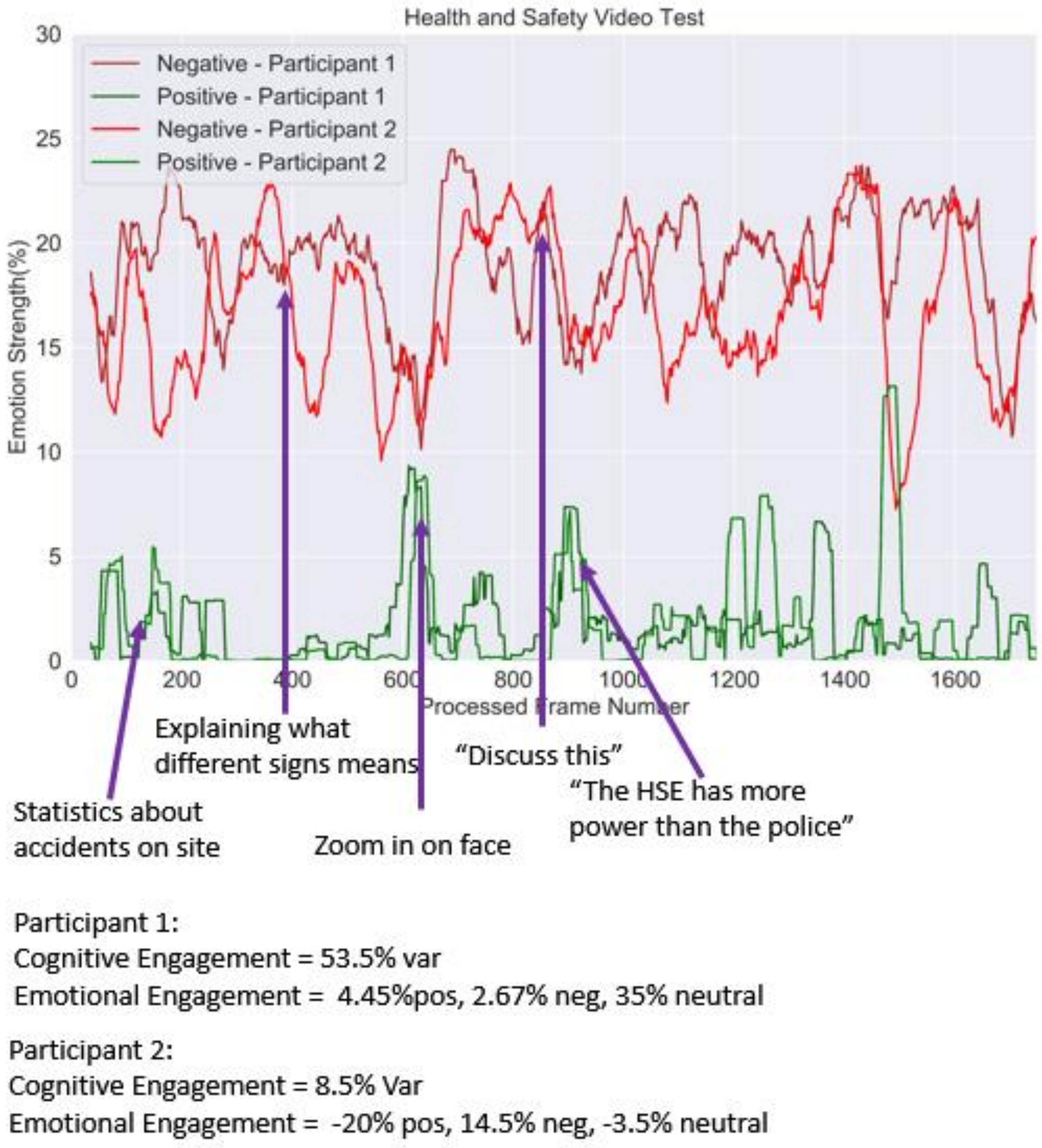

Figure 25 Positive and negative emotion of the authors throughout a pre-recorded site induction. Arrows indicate where different changes in emotion are believed to be due to a change in the content. Cognitive and emotional engagement scores are presented below

3.3.6 Application Example 2: Isolating key variables

Another useful output from this software is the ability isolate which variables contribute towards a more engaging site induction. By looking across multiple sessions we can notice trends that might not be obvious from looking at individual sessions (such as in section 3.3.5).

In Figure 26 and Figure 27, the outputs from sessions that involved some level of interaction (e.g. a Q and A session, a group task), and those that were a non-interactive session were compared for both participants. Although no statistical test was undertaken, the results suggest that interactive sessions lead to higher cognitive engagement, and a reduction in negative emotion.

On site, this kind of analysis could be used to study the effect of variables such as group size, time of day, type of content.

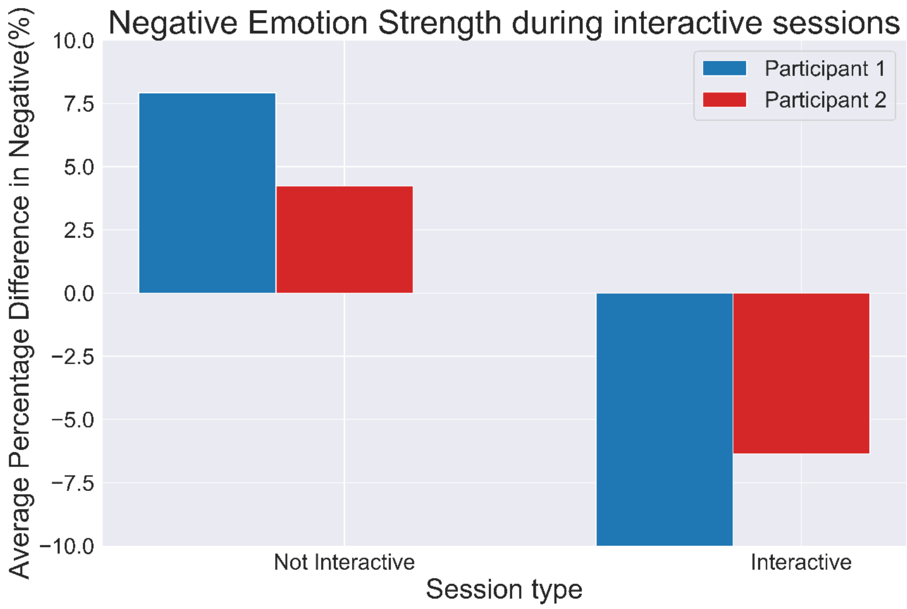

Figure 26 Average difference in negative emotion grouped by presence of interactivity for all recordings of the authors

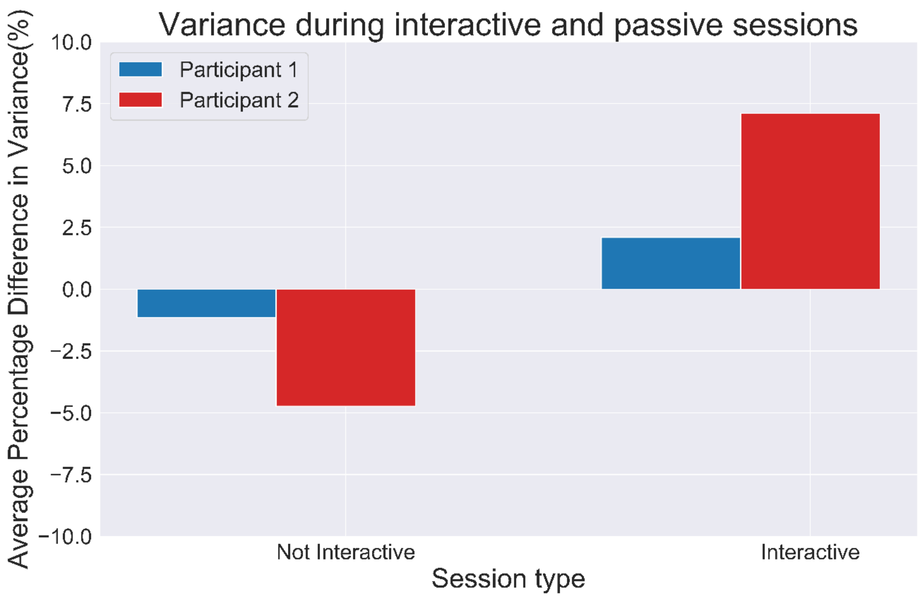

Figure 27 Average difference in variance grouped by presence of interactivity for all recordings of the authors

3.4 Program Reflections

Overall, the level of agreement between external questionnaires and FER outputs is similar to that observed in the literature (Ninaus *et al.*, 2019). It is believed the use of FER in the proposed way is an original and useful addition to the field. This study has aimed to identify a mechanism with which variables (such as gender, or type learning material) could be isolated as suggested by Ninaus *et al.*, 2019.

With further testing, it could enable for an objective understanding of exactly what variables go into making inductions successful, ultimately resulting in an improvement in safety culture and site safety.

By using facial emotion readings, the proposed method is able to successfully output three engagement scores: positive and negative emotional engagement and cognitive engagement. Whilst the preliminary use of these scores has been satisfactory, it is believed that more data points and comparison with other measures of engagement would increase the validity of this method as a tool.

In addition, the proposed metrics allow for the identification of specific instances of high engagement during sessions as well as what variables contribute to higher engagement scores throughout many sessions.

3.5 Limitations

The main identified limitation of the proposed method is the need for previous data on participants to measure ‘relative’ scores of emotional and cognitive engagement. As this method of quantifying engagement is new, few data were available for comparison. This meant that the only points of comparison for the participants were themselves and eachother. It is believed this may be impractical on some construction sites as many users of construction site only undertake a given induction once.

A possible remedy to this issue would be the gathering of large amounts of data. This would be gathered by virtue of the further testing required to verify this program on site. Then, instead of comparing an individual to their previous score, the scores of all the participants in one induction could be compared to the scores of all the participants of another induction. Assuming a normal distribution of emotivity and engagement, this would allow for similar conclusions to be drawn without the need for precedent data of each individual.

Currently, the proposed method involved the ‘manual’ analysis of the graphs for each individual in each session. Although effective, the authors believe a more efficient way of interpreting this data exists.
It was found that the most ‘events’ of a given session that triggered engagement caused spikes in the graphs. By calculating the gradients of these graphs, a python script could be written to automate the recording of when sudden changes in emotion occur- the output of which would be a list of when participants experienced a change in emotion. This would facilitate the processing of greater quantities of data.

The proposed method retains some value on a more individual basis. It has been seen in the field of personalised medicine, for example, that gaining an in-depth understanding of the needs of individuals can yield life changing outcomes (Chan and Ginsburg, 2011). In a similar way, gaining a in depth understanding of how individuals respond to site inductions could enable the development of personalised site inductions with significant improvement in outcomes.

There is a risk of introducing bias into the results by using sessions only recorded by the authors, especially in the pre-existing knowledge of emotional responses, engagement and FER systems from the literature review conducted before testing.

By nature of the testing environment differing significantly from its intended application environment, it is believed the outputs of the proposed method could show some discrepancy in outputs. Compared to a construction site setting, a domestic setting may have affected the emotional responses of the participants. In addition, the sessions studied were undertaken in an isolated environment. Although some sessions involved some interactivity, there was no physical presence of other individuals as there usually is in site inductions.

As the content the participants were watching during the recorded sessions served a different purpose to the content typically presented to site workers during inductions, it is believed a discrepancy in affect may exist. For example site workers are not pressured to learn by external motivating factors such as exams (Harvey *et al.*, 2020).

Nonetheless, there are overlaps between the purpose of lectures and site inductions in that they are both learning environments where the role of the participant is primarily to learn. To address this limitation, one session was undertaken where the authors watched a UK construction site health and safety video. The limitation of environment and isolation remain for this test, but the purpose of the session is more representative of the intended context for the FER program.

4.0 Future Work

There are myriad opportunities to progress this project further, particularly once construction sites are operating without Covid-19 restrictions in place. In the first instance, to further validate these metrics, a more in-depth validation procedure is required in the relevant context and with a larger sample size.

Following this, it is believed that the most impactful research could be undertaken by applying this method on site and testing whether changing variables such group size, time of day, type of content presentation and frequency of change in presentation style affects how engaged the workers are. This would begin to directly tackle the overarching purpose of making this program: to improve the current state of H&S outcomes on construction sites in the UK by providing objective and quantitative data.

An additional progression of this work could be to enrich the outputs of the program by incorporating more machine learning techniques such as gaze and posture recognition. There are a few examples of this currently in research projects, specifically regarding students at school and their engagement with the content (Mello, Chipman and Graesser, 2007; Zhao, Li and Jia, 2021) which suggest that posture predicts affect. This would allow further inference of a participant’s response as a result of their body language as well as their facial expressions. Aspects like nodding and slouching could be recorded and used to confirm or question the program’s suggested engagement scores.

5.0 Conclusion

This project has aimed to contribute to the field of construction health and safety research by developing a machine learning-based method of quantifying engagement during site inductions.

The primary contribution of this work is in the form of three engagement scores. Within the situational constraints of the project, these have enabled a primary analysis of how individuals respond to content. The process of isolating what variables that make an effective site induction has also been trialled. Although limited due to the amount of data the results show some promise in their potential applications.

As outlined in section 3.5, the method is preliminary and likely requires some validation and tweaking to be applicable to construction sites. Despite this, it is believed that this research lays the preliminary groundwork for providing numerical evidence in the aim of improving site inductions, and ultimately reducing accident rates on construction sites.

The potential applications of this technology to site inductions and the construction industry make it a particularly exciting topic. The benefits range from improvements in safety culture and general outlook on H&S from all involved parties, to improved site safety outcomes which would directly result in decreased injury and death rates on site.

The topics of H&S and engagement in site inductions and the potential machine learning and programming have as tools to improve the ways inductions are conducted. The main outcome of this project is a program that can output three engagement scores, allowing for further analysis of which sessions were more successful in terms of individual participant engagement. Those conducting future work could use the program to begin to identify critical components that made sessions particularly engaging. This program, after more extensive testing in the construction industry, could be a valuable tool in being able to provide numerical evidence to support a change in the structure and approach to site inductions.

6.0 GitHub Code Repository Link

https://github.com/CharlesPlusC/FER

7.0 References

Abdat, F., Maaoui, C. and Pruski, A. (2011) 'Human-computer interaction using emotion recognition from facial expression', in *Proceedings - UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, EMS 2011*, pp. 196–201. doi: 10.1109/EMS.2011.20.

Almeida, F. *et al.* (2017) 'Strengths and Limitations of Qualitative and Quantitative Research Methods Innovation and Entrepreneurship View project Observatory of Portuguese Academic Spin-offs View project European Journal of Education Studies STRENGTHS AND LIMITATIONS OF QUALITATIV', *European Journal of Education Studies*, 0(0). doi: 10.5281/zenodo.887089.

Almond, P. and Esbester, M. (2016) 'Policy and Practice in Health and Safety The changing legitimacy of health and safety, 1960-2015: understanding the past, preparing for the future'. doi: 10.1080/14773996.2016.1231868.

Amini, A. *et al.* (2019) 'Variational end-to-end navigation and localization', in *Proceedings - IEEE International Conference on Robotics and Automation*. Institute of Electrical and Electronics Engineers Inc., pp. 8958–8964. doi: 10.1109/ICRA.2019.8793579.

Andrejevic, M. and Selwyn, N. (2020) 'Facial recognition technology in schools: critical questions and concerns', *Learning, Media and Technology*, 45(2), pp. 115–128. doi: 10.1080/17439884.2020.1686014.

Arriaga, O., Plöger, P. G. and Valdenegro, M. (2017) *Real-time Convolutional Neural Networks for Emotion and Gender Classification*.

Azcarate, A. *et al.* (2005) *Automatic facial emotion recognition*.

Bagozzi, R. P., Gopinath, M. and Nyer, P. U. (1999) 'The Role of Emotions in Marketing', *Journal of the Academy of Marketing Science*, 27(2), pp. 184–206. doi: 10.1177/0092070399272005.

Bakker, A. B., Albrecht, S. L. and Leiter, M. P. (2011) 'Key questions regarding work engagement', *European Journal of Work and Organizational Psychology*, 20(1), pp. 4–28. doi: 10.1080/1359432X.2010.485352.

Baltazar, M. *et al.* (2014) 'Eye contact elicits bodily self-awareness in human adults', *Cognition*, 133(1), pp. 120–127. doi: 10.1016/j.cognition.2014.06.009.

Baxendale, T. and Jones, O. (2000) 'Construction design and management safety regulations in practice - Progress on implementation', *International Journal of Project Management*, 18(1), pp. 33–40. doi: 10.1016/S0263-7863(98)00066-0.

Beseler, C. L. and Stallones, L. (2009) 'Safety knowledge, safety behaviors, depression, and injuries in Colorado farm residents', *American Journal of Industrial Medicine*, 53(1), p. n/a-n/a. doi: 10.1002/ajim.20779.

Bontchev, B. and Vassileva, D. (2016) 'Assessing engagement in an emotionally-Adaptive applied game', in *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, pp. 747–754. doi: 10.1145/3012430.3012602.

Boschman, J. S. *et al.* (2013) 'Psychosocial work environment and mental health among construction workers', *Applied Ergonomics*, 44(5), pp. 748–755. doi: 10.1016/j.apergo.2013.01.004.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992) 'Training algorithm for optimal margin classifiers', in *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*. New York, New York, USA: Publ by ACM, pp. 144–152. doi: 10.1145/130385.130401.

Buchanan, T. W. (2007) 'Retrieval of Emotional Memories', *Psychological Bulletin*, 133(5), pp. 761–779. doi: 10.1037/0033-2909.133.5.761.

Carder, B. and Ragan, P. W. (2003) 'A survey-based system for safety measurement and improvement', *Journal of Safety Research*, 34(2), pp. 157–165. doi: 10.1016/S0022-4375(03)00007-0.

Chan, I. S. and Ginsburg, G. S. (2011) 'Personalized Medicine: Progress and Promise', *Annu. Rev. Genomics Hum. Genet. 2011*, 12, pp. 217–261. doi:

10.1146/annurev-genom-082410-101446.

Chaouachi, M. *et al.* (2010) *Affect and Mental Engagement: Towards Adaptability for Intelligent Systems*. Available at: www.aaai.org (Accessed: 27 November 2020).

Chollet, F. (2017) ‘Xception: Deep learning with depthwise separable convolutions’, in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Institute of Electrical and Electronics Engineers Inc., pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

Corder, G. and Foreman, D. (2011) *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach* | *Wiley*. Wiley. Available at: https://www.wiley.com/en-us/Nonparametric+Statistics+for+Non+Statisticians%3A+A+Step+by+Step+Approach-p-9781118165881 (Accessed: 28 March 2021).

Damen, F. (2007) *Taking the lead: The role of affect in leadership effectiveness*.

Dester, W. S. and Blockley, D. I. (1995) ‘Safety - behaviour and culture in construction’, *Engineering, Construction and Architectural Management*. MCB UP Ltd, pp. 17–26. doi: 10.1108/eb021000.

DiCarlo, J. J., Zoccolan, D. and Rust, N. C. (2012) ‘How does the brain solve visual object recognition?’, *Neuron*. NIH Public Access, pp. 415–434. doi: 10.1016/j.neuron.2012.01.010.

Doherty, R. W. (1997) ‘THE EMOTIONAL CONTAGION SCALE: A MEASURE OF INDIVIDUAL DIFFERENCES’, *Journal of Nonverbal Behaviour*, 21(2), pp. 131–154.

Donovan, R. *et al.* (2020) ‘Quantifying the Links Between Personality Sub-traits and the Basic Emotions’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Science and Business Media Deutschland GmbH, pp. 521–537. doi: 10.1007/978-3-030-58802-1_37.

Dowling, R., Lloyd, K. and Suchet-Pearson, S. (2016) ‘Qualitative methods 1’, *Progress in Human Geography*, 40(5), pp. 679–686. doi: 10.1177/0309132515596880.

Ekman, P. (1991) *Telling lies : clues to deceit in the marketplace, politics, and marriage*, *W.W. Norton*.

Ekman, P. (1999) ‘Basic Emotions’, in Dalgleish, T. and Power, M. (eds) *Handbook of Cognition and Emotion* . John Wiley and Sons.

Ekman, P. (2004) ‘Emotions revealed’, *BMJ*, 328(Suppl S5), p. 0405184. doi: 10.1136/sbmj.0405184.

Ekman, P. and Friesen, W. (1969) ‘The Repertoire of Nonverbal Behaviour: Categories, Origins, Usage, and Coding’, *Semiotica*, 1(1), pp. 49–98. Available at: https://1ammce38pkj41n8xkp1iocwe-wpengine.netdna-ssl.com/wp-content/uploads/2013/07/The-Repertoire-Of-Nonverbal-Behavior-Categories-Origins-.pdf (Accessed: 28 March 2021).

Everingham, M. *et al.* (2007) *The PASCAL Visual Object Classes Challenge 2007 (VOC2007)*. Available at: http://host.robots.ox.ac.uk/pascal/VOC/voc2007/ (Accessed: 15 November 2020).

Falchikov, N. and Boud, D. (1989) ‘Student Self-Assessment in Higher Education: A Meta-Analysis’, *Review of Educational Research*, 59(4), pp. 395–430. doi: 10.3102/00346543059004395.

Fastenrath, M. *et al.* (2014) ‘Dynamic modulation of amygdala-hippocampal connectivity by emotional arousal’, *Journal of Neuroscience*, 34(42), pp. 13935–13947. doi: 10.1523/JNEUROSCI.0786-14.2014.

Felzenszwalb, P. F. *et al.* (no date) *Object Detection with Discriminatively Trained Part Based Models*.

Fergus, R. (2015) *Neural Networks*.

Fredricks, J. A., Blumenfeld, P. C. and Paris, A. H. (2004) *School Engagement: Potential of the Concept, State of the Evidence*.

Ganotice, F. A., Datu, J. A. D. and King, R. B. (2016) ‘Which emotional profiles exhibit the best learning outcomes? A person-centered analysis of students' academic emotions', *School Psychology International*, 37(5), pp. 498–518. doi: 10.1177/0143034316660147.

Goelet, P. *et al.* (1986) 'The long and the short of long-term memory - A molecular framework', *Nature*, 322(6078), pp. 419–422. doi: 10.1038/322419a0.

Goldberg, P. *et al.* (2019) 'Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction', *Educational Psychology Review*. Springer, pp. 1–23. doi: 10.1007/s10648-019-09514-z.

Gonzalez-Sanchez, J. *et al.* (2011) 'ABE: An agent-based software architecture for a multimodal emotion recognition framework', in *Proceedings - 9th Working IEEE/IFIP Conference on Software Architecture, WICSA 2011*, pp. 187–193. doi: 10.1109/WICSA.2011.32.

Grafsgaard, J. F. *et al.* (2013) 'Automatically recognizing facial indicators of frustration: A learning-centric analysis', in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, pp. 159–165. doi: 10.1109/ACII.2013.33.

Gu, S. *et al.* (2016) 'Neuromodulator and Emotion Biomarker for Stress Induced Mental Disorders', *Neural Plasticity*. Hindawi Limited. doi: 10.1155/2016/2609128.

Guardian News '(2) Mark Zuckerberg testifies before Congress - watch live - YouTube'. Washington DC, April. Available at: https://www.youtube.com/watch?v=mZaec_mlq9M (Accessed: 31 March 2021).

Gulshan, V. *et al.* (2016) 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs Editorial Supplemental content'. doi: 10.1001/jama.2016.17216.

Hamilton, A. B. and Finley, E. P. (2019) 'Qualitative methods in implementation research: An introduction', *Psychiatry Research*, 280, p. 112516. doi: 10.1016/j.psychres.2019.112516.

Harvey, E. J. *et al.* (2020) ‘The use of actor-based immersive health and safety inductions: Lessons from the Thames Tideway Tunnel megaproject’, *Applied Ergonomics*, 82. doi: 10.1016/j.apergo.2019.102955.

Hatfield, E. *et al.* (1994) *Emotional Contagion* . Press Syndicate of the University of Cambridge.

Health and Safety Executive (2019) *Workplace fatal injuries in Great Britain 2018*, *Health and Safety Executive*. London. Available at: http://www.hse.gov.uk/statistics/pdf/fatalinjuries.pdf (Accessed: 23 November 2020).

Health and Safety Executive (2020) *Construction statistics in Great Britain*. Available at: https://www.hse.gov.uk/statistics/industry/construction.pdf (Accessed: 24 November 2020).

Hui Ma and Turgay Celik (2019) ‘FER-Net: facial expression recognition using densely connected convolutional network’, *ELECTRONICS LETTERS*, 55(4), pp. 184-–186. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8643885 (Accessed: 22 November 2020).

Isen, A. M. *et al.* (1978) ‘Affect, accessibility of material in memory, and behavior: A cognitive loop?’, *Journal of Personality and Social Psychology*, 36(1), pp. 1–12. doi: 10.1037/0022-3514.36.1.1.

Isen, A. M. (2001) ‘An Influence of Positive Affect on Decision Making in Complex Situations: Theoretical Issues With Practical Implications’, *Journal of Consumer Psychology*, 11(2), pp. 75–85. doi: 10.1207/s15327663jcp1102_01.

Jack, R. E., Garrod, O. G. B. and Schyns, P. G. (2014) ‘Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time’, *Current Biology*, 24(2), pp. 187–192. doi: 10.1016/j.cub.2013.11.064.

Kenney, E. and Adhikari, K. (2016) ‘Recent developments in identifying and quantifying emotions during food consumption’, *Journal of the Science of Food and Agriculture*, 96(11), pp. 3627–3630. doi: 10.1002/jsfa.7717.

King, S. C. and Meiselman, H. L. (2010) ‘Development of a method to measure consumer emotions associated with foods’, *Food Quality and Preference*, 21(2), pp. 168–177. doi: 10.1016/j.foodqual.2009.02.005.

Kolakowska, A. *et al.* (2013) ‘Emotion recognition and its application in software engineering’, in *2013 6th International Conference on Human System Interactions, HSI 2013*, pp. 532–539. doi: 10.1109/HSI.2013.6577877.

Langford, D., Rowlinson, S. and Sawacha, E. (2000a) ‘Safety behaviour and safety management: its influence on the attitudes of workers in the UK construction industry’, *Engineering Construction and Architectural Management*, 7(2), pp. 133–140. doi: 10.1046/j.1365-232x.2000.00137.x.

Langford, D., Rowlinson, S. and Sawacha, E. (2000b) ‘Safety behaviour and safety management: its influence on the attitudes of workers in the UK construction industry’, *Engineering Construction and Architectural Management*, 7(2), pp. 133–140. doi: 10.1046/j.1365-232x.2000.00137.x.

Lasri, I., Solh, A. R. and Belkacemi, M. El (2019) ‘Facial Emotion Recognition of Students using Convolutional Neural Network’, in *2019 3rd International Conference on Intelligent Computing in Data Sciences, ICDS 2019*. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ICDS47004.2019.8942386.

Lawani, K., Hare, B. and Cameron, I. (2017) *DEVELOPING A WORKER ENGAGEMENT MATURITY MODEL FOR IMPROVING OCCUPATIONAL SAFETY AND HEALTH (OSH) IN CONSTRUCTION*.

Leue, A. and Beauducel, A. (2011) ‘The PANAS Structure Revisited: On the Validity of a Bifactor Model in Community and Forensic Samples’, *Psychological Assessment*, 23(1), pp. 215–225. doi: 10.1037/a0021400.

Li, F.-F., Johnson, J. and Yeung, S. (2018) *Lecture 5: Convolutional Neural Networks*.

Lien, J. J. *et al.* (1998) ‘Automated facial expression recognition based on FACS action units’, in *Proceedings - 3rd IEEE International Conference on*

Automatic Face and Gesture Recognition, FG 1998. IEEE Computer Society, pp. 390–395. doi: 10.1109/AFGR.1998.670980.

Lishner, D. A., Cooter, A. B. and Zald, D. H. (2008) ‘Addressing measurement limitations in affective rating scales: Development of an empirical valence scale’, *Cognition and Emotion*, 22(1), pp. 180–192. doi: 10.1080/02699930701319139.

Lucas, R., Diener, E. and Larsen, R. (2003) ‘Positive Psychological Assessment: A Handbook of Models and Measures’, in. American Psychological Association, pp. 201–218. Available at: https://psycnet.apa.org/fulltext/2003-02181-013.pdf?auth_token=273ab52dd3ce804a2d8d3b489418a75fd56cb457 (Accessed: 25 March 2021).

Martinez-Martin, N. (2019) ‘What are important ethical implications of using facial recognition technology in health care?’, *AMA Journal of Ethics*, 21(2), pp. 180–187. doi: 10.1001/amajethics.2019.180.

MathWorks (2020) *Convolutional Neural Network - MATLAB & Simulink*, *What Makes CNNs So Useful ?* Available at: https://uk.mathworks.com/discovery/convolutional-neural-network-matlab.html (Accessed: 26 November 2020).

Meldrum, A. and Cameron, I. (2009) ‘Road testing a health and safety worker engagement tool-kit in the construction industry’. doi: 10.1108/09699980911002601.

Mello, S. D. ’, Chipman, P. and Graesser, A. (2007) ‘Posture as a Predictor of Learner’s Affective Engagement Posture as a Predictor of Learner’s Affective Engagement’, *Proceedings of the Annual Meeting of the Cognitive Science Society*, (29), p. 29.

Minaee, S. and Abdolrashidi, A. (no date) *Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network*.

Nielson, K. A. and Powless, M. (2007) ‘Positive and negative sources of emotional arousal enhance long-term word-list retention when induced as long as 30 min after learning', *Neurobiology of Learning and Memory*, 88(1), pp. 40–47. doi: 10.1016/j.nlm.2007.03.005.

Ninaus, M. *et al.* (2019a) 'Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data', *Computers and Education*, 142. doi: 10.1016/j.compedu.2019.103641.

Ninaus, M. *et al.* (2019b) 'Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data', *Computers and Education*, 142, p. 103641. doi: 10.1016/j.compedu.2019.103641.

Noah, T., Schul, Y. and Mayo, R. (2018) 'When both the original study and its failed replication are correct: Feeling observed eliminates the facial-feedback effect', *Journal of Personality and Social Psychology*, 114(5), pp. 657–664. doi: 10.1037/pspa0000121.

Ortony, A. and Turner, T. J. (1990) 'What's basic about basic emotions?', *Psychological Review*, 97(3), pp. 315–331. doi: 10.1037/0033-295X.97.3.315.

Paul Ekman, Wallace Freiesen and Sonia Ancoli (1983) 'Facial Signs of Emotional Experience', *Journal of Personal8ty and Social Psychology*, 39(6), pp. 1123–1134. Available at: https://www.paulekman.com/wp-content/uploads/2013/07/Facial-Sign-Of-Emotional-Experience.pdf (Accessed: 22 November 2020).

Pekrun, R. (2006) 'The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice', *Educational Psychology Review*, 18(4), pp. 315–341. doi: 10.1007/s10648-006-9029-9.

Pekrun, R. and Linnenbrink-Garcia, L. (2012) 'Academic emotions and student engagement', in *Handbook of Research on Student Engagement*. Springer US, pp. 259–282. doi: 10.1007/978-1-4614-2018-7_12.

Phelps, C. *et al.* (2001) *HEALTH AND SAFETY LABORATORY HSE*

Authorising Officer Norman T Byrom HSL Report Approval.

Pink, S. *et al.* (2010) ‘Building Research & Information Ethnographic methodologies for construction research: knowing, practice and interventions’. doi: 10.1080/09613218.2010.512193.

Plass, J. and Kaplan, U. (2016) *Chapter 7 - Emotional Design in Digital Media for Learning*. Elsevier Press. Available at: https://reader.elsevier.com/reader/sd/pii/B9780128018569000074?token=57BFE38830B660E0B72A02BE65F74558F90BBF2CAE477ED725241547B5C44A005A91217BF018A3971135E4E70C669AAB (Accessed: 3 March 2021).

Rotgans, J. I. and Schmidt, H. G. (2011) ‘Cognitive engagement in the problem-based learning classroom’, *Advances in Health Sciences Education*, 16(4), pp. 465–479. doi: 10.1007/s10459-011-9272-9.

Rowlinson, S. (2004) *Construction Safety Management Systems*. Spon Press. Available at: https://books.google.co.uk/books?hl=en&lr=&id=iHaBAgAAQBAJ&oi=fnd&pg=PA1&dq=injuries+on+construction+site+neglect&ots=DcUYuerP7a&sig=WeXLouLsJ_PCZG45yjSlMN2UUxA&redir_esc=y#v=onepage&q=injuries on construction site neglect&f=false (Accessed: 24 November 2020).

Rutherford, H. J. V and Lindell, A. K. (2011) ‘Thriving and Surviving: Approach and Avoidance Motivation and Lateralization’, *Emotion Review*, 3(3), pp. 333–343. doi: 10.1177/1754073911402392.

Schaufeli, W. B. and Bakker, A. B. (2004) ‘Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study’, *Journal of Organizational Behavior*, 25(3), pp. 293–315. doi: 10.1002/job.248.

Seaman, C. B. (1999) ‘Qualitative methods in empirical studies of software engineering’, *IEEE Transactions on Software Engineering*, 25(4), pp. 557–572. doi: 10.1109/32.799955.

Seltman, H. J. (2018) *Experimental Design and Analysis*. Available at: http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf (Accessed: 28 March

2021).

Serengil, S. I. and Ozpinar, A. (2020) ‘LightFace: A Hybrid Deep Face Recognition Framework’, in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, pp. 1–5. doi: 10.1109/ASYU50717.2020.9259802.

Sherratt, F. (2016) *Shiny Happy People? UK Construction Industry Health: Priorities, Practice and Public Relations*, *Association of Researchers in Construction Management*.

Sherratt, F., Farrell, P. and Noble, R. (2012) ‘Construction Management and Economics UK construction site safety: discourses of enforcement and engagement UK construction site safety: discourses of enforcement and engagement’. doi: 10.1080/01446193.2012.747689.

Shouse, E. (2005) ‘Feeling, Emotion, Affect ’, *M/C Journal*, 8(6). Available at: https://journal.media-culture.org.au/index.php/mcjournal/article/view/2443 (Accessed: 31 March 2021).

Spillane, J. and Oyedele, L. (2013) ‘Strategies for effective management of health and safety in confined site construction’, *Australasian Journal of Construction Economics and Building*, 13(4), pp. 50–64. doi: 10.5130/ajceb.v13i4.3619.

Spinelli, S. *et al.* (2014) ‘How does it make you feel? A new approach to measuring emotions in food product experience’, *Food Quality and Preference*, 37, pp. 109–122. doi: 10.1016/j.foodqual.2013.11.009.

Susskind, J. M. *et al.* (2007) ‘Human and computer recognition of facial expressions of emotion’, *Neuropsychologia*, 45(1), pp. 152–162. doi: 10.1016/j.neuropsychologia.2006.05.001.

Tadas Baltrusaitis *et al.* (2018) ‘OpenFace 2.0: Facial Behavior Analysis Toolkit’, in *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 59–66. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8373812

(Accessed: 22 November 2020).

Taigman, Y. *et al.* (2014) *DeepFace: Closing the gap to human-level performance in face verification*, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2014.220.

Teixeira, T., Wedel, M. and Pieters, R. (2012) ‘Emotion-Induced Engagement in Internet Video Advertisements’, *Journal of Marketing Research*, 49(2), pp. 144–159. doi: 10.1509/jmr.10.0207.

Thompson, E. R. (2007) ‘Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS)’, *Journal of Cross-Cultural Psychology*, 38(2), pp. 227–242. doi: 10.1177/0022022106297301.

Vanhamme, J. (2000) ‘The Link Between Surprise and Satisfaction: An Exploratory Research on how best to Measure Surprise’, *Journal of Marketing Management*, 16(6), pp. 565–582. doi: 10.1362/026725700785045949.

Watson, D., Lee Anna, C. and 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), pp. 1063–1070.

Whitehill, J. *et al.* (2014) ‘The faces of engagement: Automatic recognition of student engagement from facial expressions’, *IEEE Transactions on Affective Computing*, 5(1), pp. 86–98. doi: 10.1109/TAFFC.2014.2316163.

Whiteoak, J. W. and Mohamed, S. (2016) ‘Employee engagement, boredom and frontline construction workers feeling safe in their workplace’, *Accident Analysis and Prevention*, 93, pp. 291–298. doi: 10.1016/j.aap.2015.11.001.

Wieckowski, A. T. and White, S. W. (2017) ‘Eye-Gaze Analysis of Facial Emotion Recognition and Expression in Adolescents with ASD’, *Journal of Clinical Child and Adolescent Psychology*, 46(1), pp. 110–124. doi: 10.1080/15374416.2016.1204924.

WJEC Construction (2016) *Health and Safety tutorial [YouTube]*. Available at:

https://www.youtube.com/watch?v=eW7GBuaBuek (Accessed: 29 March 2021).

Xu, F., Zhang, J. and Wang, J. Z. (2017) ‘Microexpression Identification and Categorization Using a Facial Dynamics Map’, *IEEE Transactions on Affective Computing*, 8(2), pp. 254–267. doi: 10.1109/TAFFC.2016.2518162.

Yu, J. *et al.* (2015) ‘Being watched by others eliminates the effect of emotional arousal on inhibitory control’, *Frontiers in Psychology*, 6(JAN). doi: 10.3389/fpsyg.2015.00004.

Zhao, J., Li, J. and Jia, J. (2021) ‘A study on posture-based teacher-student behavioral engagement pattern’, *Sustainable Cities and Society*, 67, p. 102749. doi: 10.1016/j.scs.2021.102749.

Zhao, M., Adib, F. and Katabi, D. (2016) ‘Emotion recognition using wireless signals’, in *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*. New York, NY, USA: Association for Computing Machinery, pp. 95–108. doi: 10.1145/2973750.2973762.

Zheng, Z. *et al.* (2016) ‘Safety Needs Mediate Stressful Events Induced Mental Disorders’, *Neural Plasticity*. Hindawi Limited. doi: 10.1155/2016/8058093.

Zull, J. E. (2006) ‘Key aspects of how the brain learns’, *New Directions for Adult and Continuing Education*, 2006(110), pp. 3–9. doi: 10.1002/ace.213.

8.0 Appendices

8.1 Appendix A – PANAS and Engagement Questionnaire Forms Used

	Very slightly or not at all	**A little**	**Moderately**	**Quite a bit**	**Extremely**
Interested	☐	☐	☐	☐	☐
Distressed	☐	☐	☐	☐	☐
Excited	☐	☐	☐	☐	☐

Upset	☐	☐	☐	☐	☐
Strong	☐	☐	☐	☐	☐
Guilty	☐	☐	☐	☐	☐
Scared	☐	☐	☐	☐	☐
Hostile	☐	☐	☐	☐	☐
Enthusiastic	☐	☐	☐	☐	☐
Proud	☐	☐	☐	☐	☐
Irritable	☐	☐	☐	☐	☐
Alert	☐	☐	☐	☐	☐
Ashamed	☐	☐	☐	☐	☐
Inspired	☐	☐	☐	☐	☐
Nervous	☐	☐	☐	☐	☐
Determined	☐	☐	☐	☐	☐
Attentive	☐	☐	☐	☐	☐
Jittery	☐	☐	☐	☐	☐
Active	☐	☐	☐	☐	☐
Afraid	☐	☐	☐	☐	☐

PANAS Questionnaire sheet was created from the work of Watson, Lee Anna and Tellegen (1988).

Cognitive Engagement Questionnaire

	Not True at all for me	**Not True for me**	**Neutral**	**True for me**	**Very true for me**

I put in a lot of effort	☐	☐	☐	☐	☐
I was engaged with the topic at hand and I wish we could still continue with the session for a while	☐	☐	☐	☐	☐
I was so involved that I forgot everything around me	☐	☐	☐	☐	☐

The cognitive engagement questionnaire sheet was created from the work of Rotgans and Schmidt (2011).

8.2 Appendix B – PANAS and Engagement Questionnaire Results

PANAS – Participant 2																	
Session date	12.1.21	12.1.21	13.1.21	13.1.21	15.1.21	15.1.21	18.1.21	19.1.21	19.1.21	21.1.21	21.1.21	25.1.21	25.1.21	26.1.21	28.1.21	1.2.21	2.2.21
Session No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Time (24hr)	11	15	9	10	14	15	14	10	14	16	17	14	15	14	18	11	14
Interested	3	4	4	4	4	5	2	1	4	3	5	1	1	3	2	3	3
Distressed	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
Excited	1	3	2	2	3	2	1	1	2	2	4	1	1	1	1	1	1
Upset	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
Strong	1	2	1	1	2	1	1	1	1	2	1	1	1	2	1	1	2

Guilty	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
scared	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Hostile	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Enthusiastic	1	4	3	4	3	4	2	1	2	3	5	1	1	2	1	2	3
Proud	1	2	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1
Irritable	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Alert	2	5	4	5	5	3	2	1	4	3	5	2	2	4	4	3	4
Ashamed	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Inspired	1	4	2	3	5	3	1	1	3	4	5	1	1	2	1	2	2
Nervous	1	2	1	1	2	2	1	1	2	1	2	1	1	1	2	2	2
Determined	1	2	2	3	2	2	1	1	3	1	4	1	3	2	1	4	2

Attentive	2	2	3	3	3	3	2	1	4	3	4	2	2	4	2	4	2
Jittery	2	2	1	2	2	2	1	1	2	1	2	1	1	1	1	1	1
Active	1	2	1	4	3	2	2	1	3	2	5	1	1	2	1	1	2
Afraid	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Positive Affect Score	14	30	23	30	31	26	15	10	27	25	40	12	14	23	15	22	22
Negative Affect Score	11	12	10	11	12	12	10	10	12	10	13	10	10	10	12	11	11

Positive Decimal	0.1	0.5	0.325	0.5	0.525	0.4	0.125	0	0.425	0.375	0.75	0.05	0.1	0.325	0.125	0.3	0.3
Negative Decimal	0.025	0.05	0	0.025	0.05	0.05	0	0	0.05	0	0.075	0	0	0	0.05	0.0	0.0

																2 5	2 5

	PANAS - Participant 1										
	Session No.	1	2	3	4	5	6	7	8	9	10
	Time (24hr)	14	10	16	14	11	9	10	11	14	16
1	Interested	3	4	2	1	2	3	3	1	3	3
2	Distressed	1	1	1	1	1	2	2	3	3	1
3	Excited	2	3	1	1	1	3	3	2	2	2
4	Upset	3	1	4	2	3	1	1	3	4	4
5	Strong	3	3	3	3	2	3	3	2	2	2
6	Guilty	1	1	1	1	1	1	1	2	2	1
7	scared	1	1	2	1	1	2	2	2	2	2
8	Hostile	3	1	3	2	3	1	1	1	1	1

9	Enthusiastic	3	4	1	2	2	3	3	2	2	2
10	Proud	2	3	2	2	2	3	3	2	2	2
11	Irritable	3	2	3	3	3	1	1	1	1	1
12	Alert	3	3	2	1	2	2	2	2	2	2
13	Ashamed	1	1	1	1	1	1	1	1	1	1
14	Inspired	1	4	1	1	1	2	2	1	1	1
15	Nervous	1	2	1	1	2	1	1	1	1	1
16	Determined	3	3	2	1	1	3	3	2	2	2
17	Attentive	3	4	2	2	1	3	2	1	1	1
18	Jittery	2	1	2	3	3	3	3	3	3	1
19	Active	2	3	3	2	2	3	2	1	1	1
20	Afraid	1	1	1	1	1	1	1	1	1	1

Positive	25	34	19	16	16	28	26	16	18	18

Negative	17	12	19	16	19	14	14	18	19	14

Positive Decimal	0.375	0.6	0.225	0.15	0.15	0.45	0.4	0.15	0.2	0.2
Negative Decimal	0.175	0.05	0.225	0.15	0.225	0.1	0.1	0.2	0.225	0.1

Table 4 The Cognitive Engagement Scores for each Session

Engagement Questionnaire - Participant 2																	
Session No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Time (24hr)	11	15	9	10	14	15	14	10	14	16	17	14	15	14	18	11	14
I was engaged with the topic at hand	2	1	3	2	1	2	2	3	3	1	2	4	3	2	1	2	3
‘I put in a lot of effort;‘I wish we could still continue with the work for a while’	3	1	2	3	1	3	2	2	4	3	3	4	3	1	2	2	1
I was so involved that I forgot everything around me	4	2	2	2	1	2	2	2	3	1	3	3	2	1	3	3	2

Total Score	3	1	2	2	1	2	2	2	3	2	3	4	3	1	2	2	2
Rounded to nearest 0.5		2	3	3	1	3	2	3	4	2	3	4	3	2	2	3	2

Engagement Questionnaire - Participant 1										
Session No.	1	2	3	4	5	6	7	8	9	10
Time (24hr)	14	10	16	14	11	9	10	11	14	16
I was engaged with the topic at hand	2	1	3	2	2	3	2	3	1	1
‘I put in a lot of effort;‘I wish we could still continue with the work for a while’	2	1	4	3	3	3	3	3	2	3
I was so involved that I forgot everything around me	3	2	3	2	2	3	2	3	1	2
Total Score	2	1	3	2	2	3	2	3	1	2

Rounded to nearest 0.5	2	2	4	3	3	3	3	3	2	2

8.3 Appendix C – Positive and Negative Emotion and Variance scores

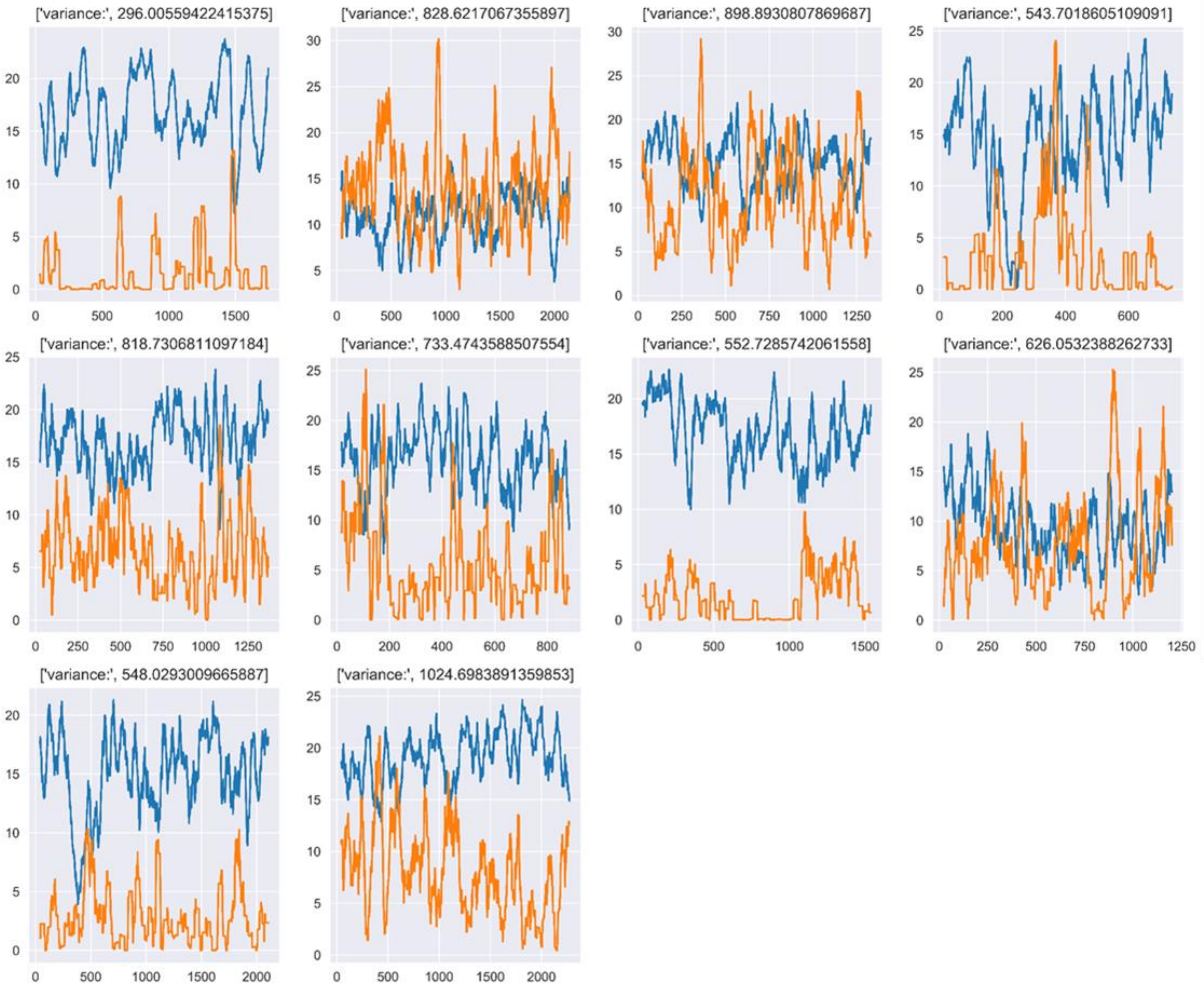

Figure 28 Positive and negative emotion, with corresponding variance score of all recordings of participant 1

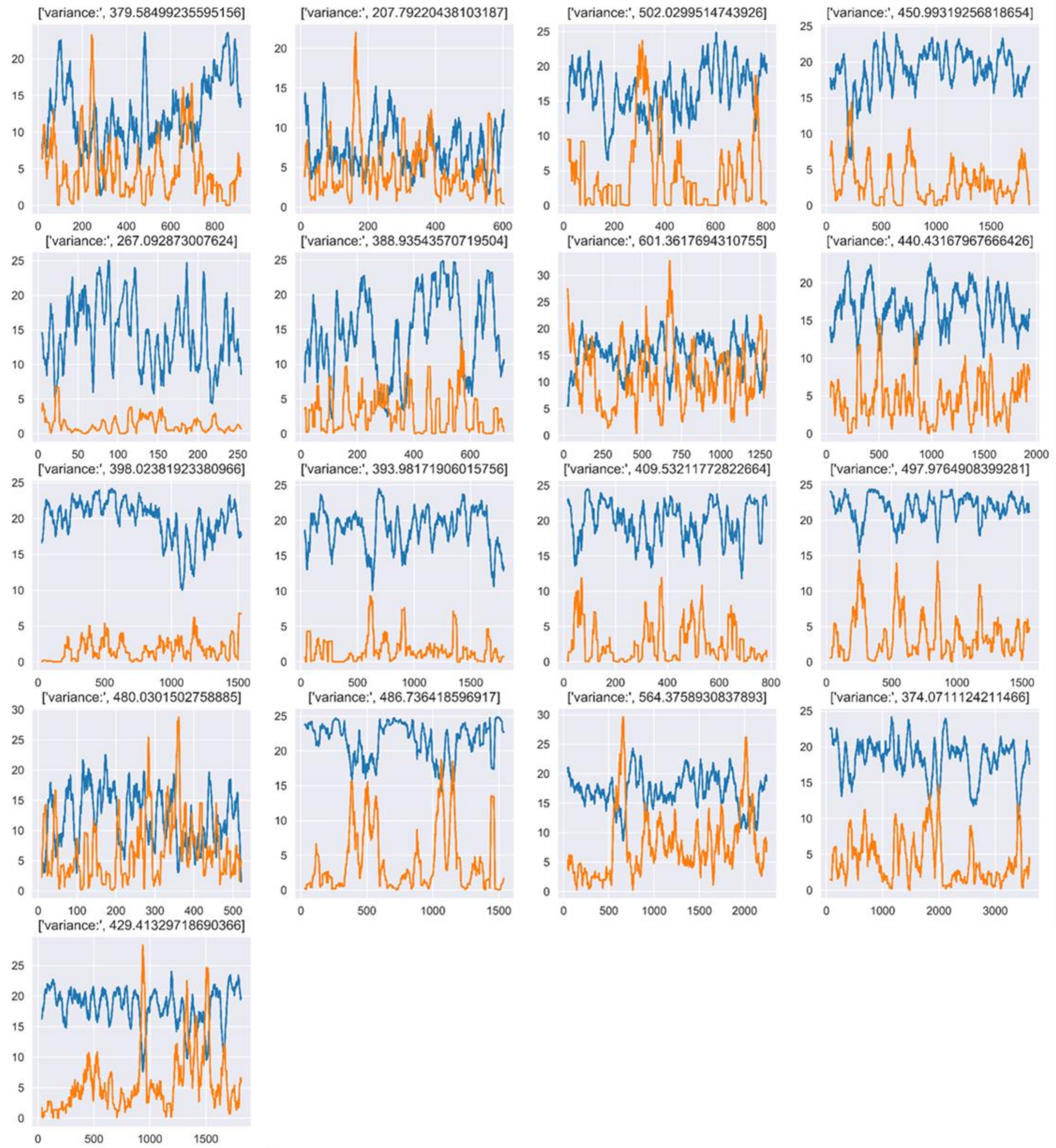

Figure 29 Positive and negative emotion, with corresponding variance score of all recordings of participant2