Automated Facial Expression Detection
An improved method for assessing engagement with computer-based training

John Fairchild

SAIC

Orlando, Florida

john.r.fairchild@saic.com

Shawn Stafford,
Adams Greenwood-Ericksen

Full Sail University

Winter Park, Florida

sstafford@fullsail.com,
agreenwooderickson@fullsail.com

ABSTRACT

Two studies were performed to assess the effectiveness of a newly developed automated facial expression detection tool in assessing the quality of trainee engagement with training material relative to traditional Kirkpatrick level-1 style assessment survey techniques. The first study (N=17) established broad support for the superiority of automated facial expression recognition tools compared to traditional techniques. A follow up study (N = 37) was performed to refine and extend findings from the first study. Taken together, the two studies showed the superiority of automated facial expression detection in volume of data collected, identification of key performance indicators, and suitability for distributed online and other computer based learning environments. Both experiments were facilitated on consumer spec PC laptops using standard built in web camera hardware to collect participant facial expression feedback in reaction to viewing video based training content. Results show promise in future implementation regarding improvements to quantity, quality and method for data collection and analysis of user engagement with digital training.

ABOUT THE AUTHORS

John Fairchild is a Solutions Architect for SAIC. With a background in serious games and a degree in game design and development, his primary efforts over the past decade have been utilizing emerging technology for training, education and outreach. Through a recent partnership with the Full Sail University Research Institute, John has worked on several IR&D efforts aimed at improving digital training methods, capabilities and standards using design theory, gamification and quest based learning, augmented and virtual reality and other game-based technology.

Dr. Stafford has an appointment at Full Sail University as the Director of Research for the University (Office of Research) and an additional appointment in the GDMS program as director of the User Experience Center. He teaches and mentors M.S. and Ph.D. students in applied video game testing, weapons studies, forensics, sensation & perception, and physiological psychology, usability, game usability. Dr. Stafford is also the Senior Research Scientist for Wargaming.net where he runs user experience studies in the User Experience Center.

Adams Greenwood-Ericksen is a Course Director at Full Sail University and is the Chief Scientist of the Full Sail User Experience Laboratory. In his role at the User Experience Lab, he has overseen over 100 commercial usability projects, as well as numerous academic research studies.

Automated Facial Expression Detection
An improved method for assessing engagement with computer-based training

John Fairchild

SAIC

Orlando, Florida

john.r.fairchild@saic.com

Shawn Stafford,
Adams Greenwood-Ericksen

Full Sail University

Winter Park, Florida

sstafford@fullsail.com,
agreenwooderickson@fullsail.com

INTRODUCTION

With the advent of digital training and other technological approaches to e-learning, including remotely distributed and online training, a need and opportunity to improve feedback collection and analysis have presented themselves. This paper presents research into a solution that utilizes automated facial feedback collection and analysis using commonly available digital cameras or webcams. Traditional assessments of training engagement and feedback suffer from limitations in quality and usefulness for a variety of reasons including lack of effort or participation, primacy recency effect, recall, and honesty.

The approach used in these studies was intended to automate the process of collecting data and feedback regarding engagement and user reaction, taking the burden off of the users/observers to supply adequate amounts of valid feedback. By leveraging involuntary micro facial responses to stimuli that are almost universally common across age, race and culture, objective real time assessments of engagement and user interest can be generated. The "7% Rule" (Mehrabian & Weiner, 1967) argues that communication is only 7% verbal and 93% non-verbal, where the non-verbal is 55% body language and facial expression. By monitoring and analyzing facial responses during consumption of digital training coursework, it is possible to tap into a considerable amount of previously underutilized data. Our results suggest that these data can be more accurate, honest, and comprehensive compared to subjective feedback collected through traditional subjective or declarative post learning surveys.

The facial recognition technology is based on a functional program built around an algorithm developed by the Media Lab at the Massachusetts Institute of Technology which utilizes a data repository of more than 3.9 million analyzed faces drawn from from over 75 countries, effectively yielding over 40 billion emotion data points. Computer vision algorithms identify key landmarks on the face, such as the tip of the nose, the edges of the eyes and the corners of the mouth. Machine learning algorithms analyze pixel data in those regions to classify facial expressions and combinations of those facial expressions are then mapped to emotions. The real-time emotion data is processed and stored for later analysis of aggregate trends across polled users (Affectiva, 2016). The results quickly highlight the times and conditions during which users are most or least engaged, become bored or contemptuous, appear happy or to be enjoying the experience, and even identifies those times when they stop paying attention all together. By noting to which elements of training users were attending when these trends in emotional response occurred, developers and instructional designers can more pragmatically identify where training is meeting learning objectives or falling short of desired results. The results show that the overall goal of providing a tool that can autonomously measure human response to digital content in an effort to improve future revisions and methods for training has largely been met.

LITERATURE AND BACKGROUND

"A growing interest into training services solutions to contain live and virtual training expenditure is expected to reach a cumulated market size of $20.39 billion during 2014–2023. Live simulated training solutions will also record a strong interest among customers to improve their actual training performance. However, computer-based training (CBT) solutions will offer the biggest stream of accessible revenues during the forecast period, accounting for $24.15 billion." (Frost & Sullivan, 2015).

Trends in DoD and private sector training show increasing reliance on digital, distributed and online training. The question: how can we better assess and improve training leveraging the same hardware used to facilitate the coursework without adding significant additional time or cost requirements. Typically, the only data collected used to evaluate and/or improve the coursework are the results of tests on the subject matter. Test data does not take into account pre-existing knowledge or other factors that can skew derived conclusions. For example, typical analysis cannot distinguish between a "lucky guess" and a "confident answer."

The other primary source of data for assessments are post course evaluations taking the form of a Kirkpatrick level 1 assessments and often optional or voluntary. A potential problem arises from this source of data when the training or assessment are not delivered in an environment where an instructor, facilitator, instructional designer or content developer can observe how the user interacted with the coursework. Post course evaluations are only as good as the effort put forth toward completing them. Often, when optional, most participants will simply skip the survey or take the path of least resistance filling in inaccurate data to fulfil requirements mandated by the facilitation.

We propose, involuntary (micro) facial reactions expressed while interacting with digital training is more prevalent and accurate than subjective post training surveys and assessments. We argue data collected in this method is not dependent on effort, recall, accuracy or even honesty on the user's part. In two research studies performed, we continuously polled the user's facial responses (multiple times per second) to capture even the slightest wince, brow furrow, smirk or smile while the subject watched informational/training videos.

The specific hypotheses addressed in these two experiments were as follows:

Hypothesis 1
Facial feedback analysis will capture all data that is captured in post analysis Kirkpatrick Level 1 questions.

Hypothesis 2
Facial expression data capture will not be affected by primacy or recency effects, whereas Kirkpatrick Level 1 Feedback will demonstrate both primacy and recency effects.

Hypothesis 3
Facial feedback analysis will capture more key moments than post session Kirkpatrick Level 1 feedback.

METHODOLOGY

Study 1

17 participants watched a series of videos that had an educational purpose. Each participant was given identical instructions, then allowed to watch the videos. While watching the videos, participants' facial expressions were recorded and analyzed using a facial data capture tool. After finishing each video, participants were directed to take a short survey. They then repeated this process for the remaining three videos. All participants were randomly chosen to partake in this study. Once all the participants' data was gathered, three lab researchers graded their qualitative responses. To grade this a rubric was created. The rubric evaluated amount of information, grammar, quality or response, and idea organization.

To increase power of the study a within subjects design was used. Video presentation order was varied to control for order effects. Each participant was given 1 video per day for 4 days to reduce interference, stress & experimental fatigue. All participants signed an informed consent and were compensated for the research study.

Based on results from study 1, an additional study was designed to investigate hypotheses 2 and 3 more closely.

Study 2

37 participants watched one of two videos (Alpha and Bravo) that had an educational purpose (teaching observers about CPR). Each participant was given identical instructions, then allowed to watch the videos. While watching the videos, participants' facial expressions were recorded and analyzed using a facial feedback data capture and analysis

tool. After finishing each video, participants were directed to take a short survey. All participants were randomly chosen to partake in this study. Once all the participants' data was gathered, two lab researchers graded their qualitative responses to assess user engagement with the material. To ensure consistency of grading, a rubric was created. The rubric evaluated the amount of information, grammar, quality and idea organization in each response.

RESULTS

Study 1

This study revealed moderate evidence of a correlation between valence, engagement, and quality of responses from Kirkpatrick level 1 qualitative feedback, as shown in the next 3 main findings:

Hypothesis 1
Facial feedback analysis will capture all data that is captured in post analysis Kirkpatrick Level 1 questions

Finding 1
In almost every case, facial capture data consistently demonstrated clear deflections around any event in the videos that participants subsequently mentioned. Figure 1 below shows aggregate facial capture data around three specifically identified events. Note the exceptionally strong correlation between recalled events and points of interest as identified by captured facial expression data.

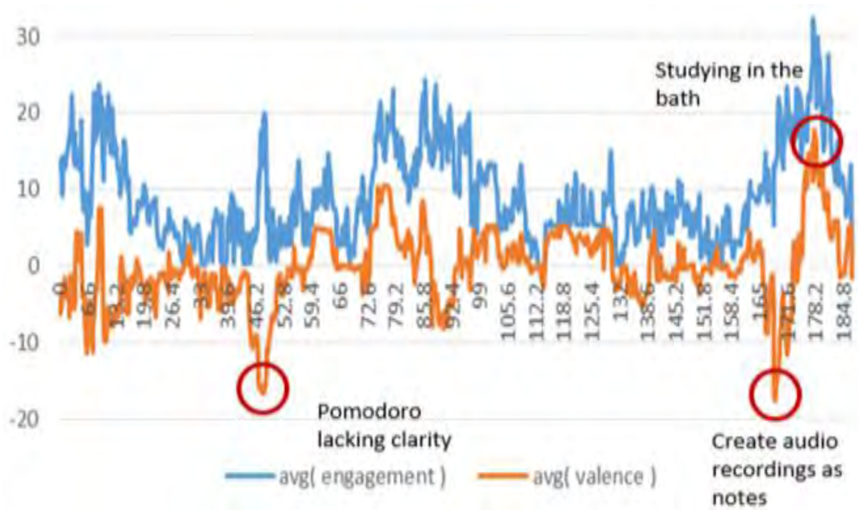

Figure 1

Hypothesis 2
Facial expression data capture will not be affected by primacy or recency effects, whereas Kirkpatrick Level 1 Feedback will demonstrate both primacy and recency effects.

Finding 2
Across all feedback, only 1 point of common Kirkpatrick feedback was outside the initial 33% of a video or the last 33% of a video (see Figure 2 below). The majority of post session qualitative feedback from participants focused on initial components of training or the final components of training. Facial feedback analysis effectively captured almost all interesting (facial responsive) points including those mentioned by participants on the Kirkpatrick Level 1 assessment, and also captured moments between and within the primacy and recency effect. In fact, in many cases it was possible to cue participants to recall and discuss unreported events by asking them about events where they had showed local peaks in facial expression capture data. This was particularly noteworthy in events which occurred in the middle of videos, where the primacy/recency effect had the most pronounced negative effects on attention and recall.

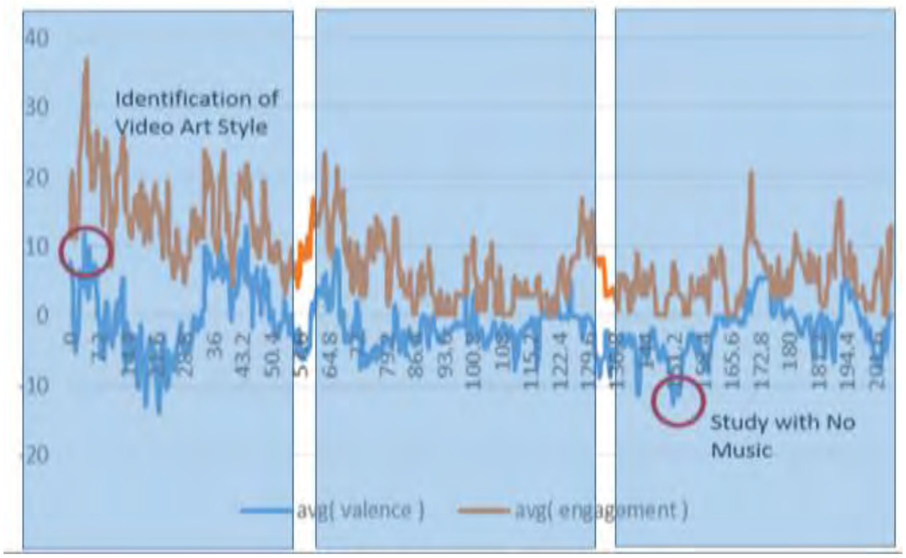

Figure 2

Hypothesis 3
Facial feedback analysis will capture more key moments than post session Kirkpatrick Level 1 feedback

Finding 3
Facial feedback analysis was observed to capture approximately 70% more key moments of response than traditional post session feedback. Furthermore, virtually all of the events captured by posttest subjective self-reporting were also identified using facial expression capture. In Figure 3 below, one example of this effect is shown. Facial capture data identified over 7 points of interest, while self-reporting produced only 2 recalled events.

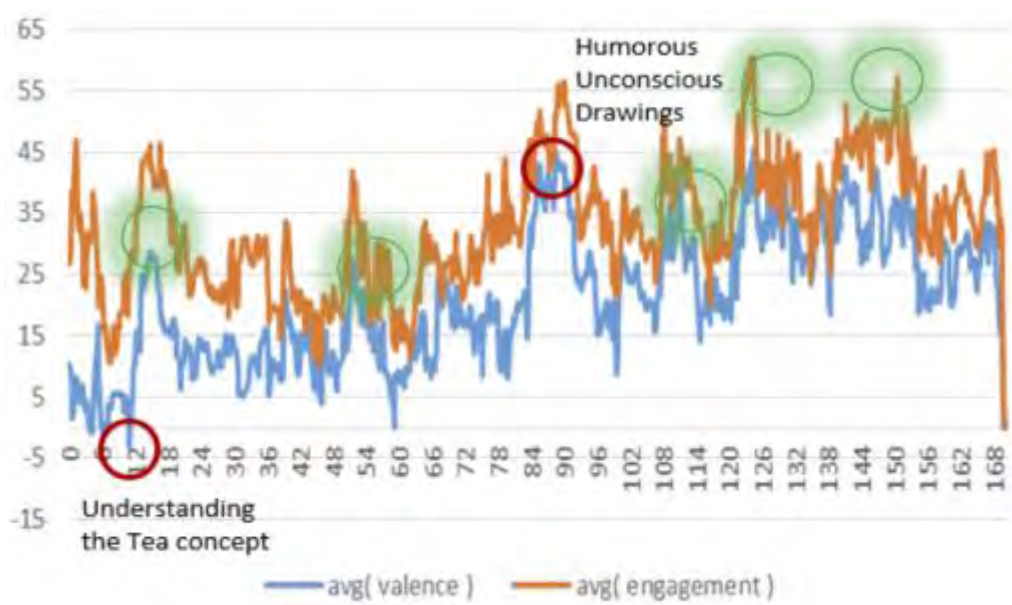

Figure 3

Study 2
A second study was conducted to replicate and extend the findings for hypotheses 2 and 3.

Hypothesis 2
Facial expression data capture will not be affected by primacy or recency effects, whereas Kirkpatrick Level 1 Feedback will demonstrate both primacy and recency effects.

To test this hypothesis further, a larger sample size was collected using longer videos to more closely assess the potential for a primacy or recency effect. In figure 4 below, points A-M indicate points where facial expression capture identified local maxima or minima indicating engagement with the learning material across all participants interacting with condition Alpha.

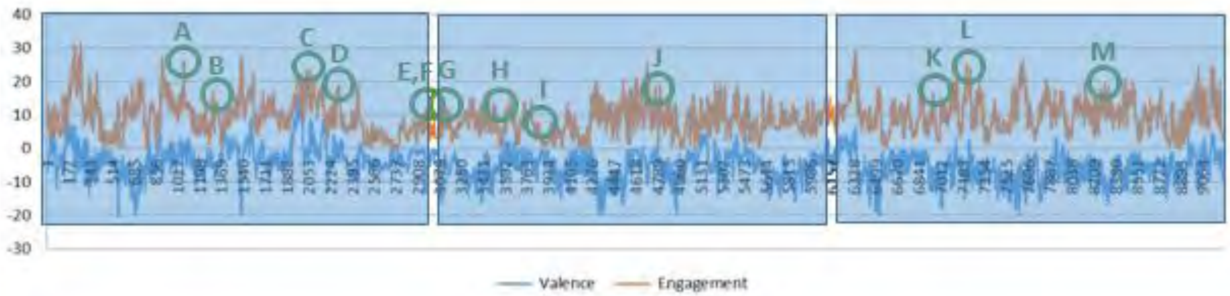

Figure 4

In Figure 5 below, points A-U represent points where facial expression capture identified local maxima or minima indicating engagement with the learning material across all participants interacting with condition Bravo.

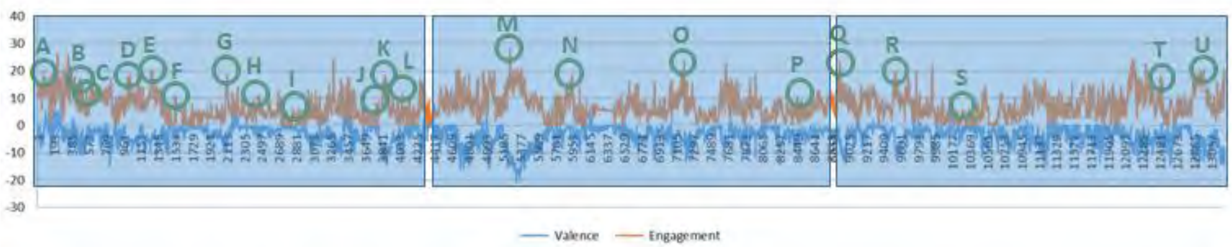

Figure 5

In Table 1 below, the means and standard deviations for the first third, second third, and last third of Alpha.

Condition Alpha						
	First Third (A1)		Second Third (A2)		Last Third (A3)	
	Average A1	SD A1	Average A2	SD A2	Average A3	SD A3
Mean	10.01	5.74	8.21	3.98	10.57	5.01
StDev	-2.85	4.58	-5.86	4.12	-6.19	4.17

Table 1

In Table 2 below, the means and standard deviations for the first third, second third, and last third of Bravo.

Condition Bravo						
	First Third (B1)		Second Third (B2)		Last Third (B3)	
	Average B1	SD B1	Average B2	SD B2	Average B3	SD B3
Mean	6.68	4.70	7.12	4.70	7.42	4.66
StDev	-2.95	3.07	-4.02	3.74	-3.66	3.19

Table 2

Finding 2

The results show both a clear lack of primacy or recency effect on facial capture data for conditions Alpha and Bravo and a slight primacy/recency effect on subjective survey data for condition Bravo. A primacy/recency effect was not observed on subjective data in condition Alpha, but this is likely due to the short duration of that condition relative to Bravo (Alpha was only 9 minutes long, whereas Bravo was 17). Generally, these results bear out the predictions of Hypothesis 2, and largely replicate the findings from Study 1.

Hypothesis 3

Facial feedback analysis will capture more key moments than post session Kirkpatrick Level 1 feedback

In Figure 6 below, points A-M indicate points where facial expression capture identified local maxima or minima indicating engagement with the learning material across all participants interacting with condition Alpha.

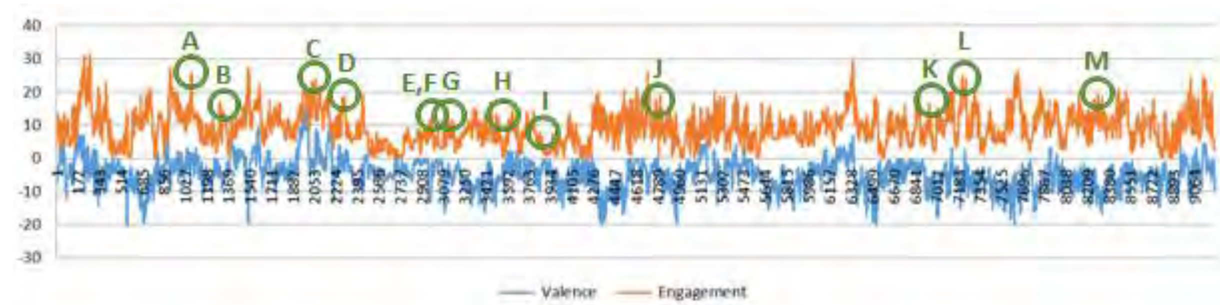

Figure 6

In Figure 7 below, points A-U represent points where facial expression capture identified local maxima or minima indicating engagement with the learning material across all participants interacting with condition Bravo.

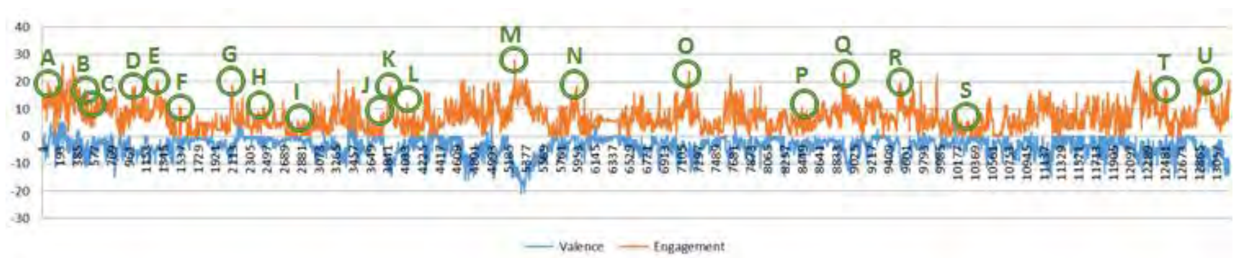

Figure 7

In table 3 below, a list of the relevant learning points for the video used in condition Alpha, and engagement measures from a traditional Kirkpatrick level 1 instrument and the facial expression capture tool.

Condition Alpha				
#	**Learning Item**	**Participants**	**Kirkpatrick**	**Facial Capture**
1	Decide to help, check for danger	3	17%	A
2	Try to get a response	0	0%	B
3	Steps to open airway	1	6%	C
4	Check for breathing	4	22%	D
5	Call Ambulance	3	17%	E
6	AED	4	22%	F
7	How do I find a defibrillator	5	28%	G
8	Where to press & hand position for CPR	5	28%	H
9	Rate, how hard and depth of CPR	6	33%	I
10	Mouth to mouth process	1	6%	J
11	Prepare subject for defibrillator	1	6%	K
12	Prepare to use the defibrillator	1	6%	L
13	Continue until Paramedics Arrive	0	0%	M

Table 3

In table 4 below, a list of the relevant learning points for the video used in condition Bravo, and engagement measures from a traditional Kirkpatrick level 1 instrument and the facial expression capture tool.

Condition Bravo				
#	**Learning Item**	**Participants**	**Kirkpatrick**	**Facial Capture**
1	Check for danger	2	11%	A
2	Try to get a response	1	5%	B
3	Call 911	1	5%	C
4	Intro to ABC check	9	47%	D
5	Check airway	4	21%	E
6	Check breathing	4	21%	F
7	Giving breaths	5	26%	G

8	Check circulation	4	21%	H
9	Check pulse	2	11%	I
10	Jaw thrust	1	5%	J
11	Breathing barrier	1	5%	K
12	Recovery position	5	26%	L
13	When do do rescue breathing	1	5%	M
14	How to do rescue breathing	0	0%	N
15	16% Oxygen, 8% Carbon Dioxide	0	0%	O
16	When to do CPR	2	11%	P
17	Signs of no blood circulation	1	5%	Q
18	How to do CPR	9	47%	R
19	CPR cycle intervals	5	26%	S
20	Working with a partner	4	21%	T
21	Defibrillation	1	5%	U

Table 4

Finding 3

For both the Alpha and Bravo training conditions, many more points of interest (characterized by significant deflections in affect and valence identified by facial capture) were observed than were identified by a Kirkpatrick level 1 style posttest survey (items recalled by at least 20% of participants). For condition Alpha, 13 data points were identified by facial capture compared to only 5 data points identified by the survey. For condition Bravo, 21 data points were identified by facial capture compared to only 9 data points identified by the survey.

SUMMARY OF RESULTS

Hypothesis 1

Facial feedback analysis will capture all data that is captured in post analysis Kirkpatrick Level 1 questions.

This hypothesis was supported by the results of Study 1.

Hypothesis 2

Facial expression data capture will not be affected by primacy or recency effects, whereas Kirkpatrick Level 1 Feedback will demonstrate both primacy and recency effects.

This hypothesis was supported by the results of both Study 1 and Study 2.

Hypothesis 3

Facial feedback analysis will capture more key moments than post session Kirkpatrick Level 1 feedback.

This hypothesis was supported by the results of both Study 1 and Study 2.

DISCUSSION

Given the strength and consistency of the findings, it seems reasonable to suggest that automated facial detection technology provides a low cost, minimally intrusive, high performance alternative to traditional Kirkpatrick level 1 survey methods when assessing training engagement and trainee affect. These techniques exceeded Kirkpatrick level 1 measures of attention and engagement, while reducing assessment time and eliminating the primacy and recency effects, a common source of uncontrolled error in assessing training engagement. Moreover, the unobtrusive nature and suitability of these methods to online distributed learning environments makes them an obvious choice for web-based and/or remote e-learning training activities.

The findings related to each hypothesis tend to bear this out. With respect to hypothesis 1, the abundance of facial expression data relative to survey style data, as well as the consistency with which collected facial expression data cover all of the points identified by the Kirkpatrick level 1 approach strongly suggest that for analyses of training engagement, this technique is an across the board improvement over traditional survey methods. However, these results should not be seen as supporting the full replacement of traditional Kirkpatrick level 1 type methods, especially since facial expression capture data provides only minimal insight into certain types of important questions, such as when assessing understanding of core content and expected reliance or trust in training content accuracy.

Additionally, the results of testing on hypothesis 2 supports the contention that no primacy or recency effect was observed in the analysis of automated facial expression capture data, whereas a slight primacy/recency effect was observed in the data obtained through traditional survey methods for condition Bravo. No effect was observed in condition Alpha, which was shorter in duration, and therefore less prone to primacy or recency effects. These findings underline a key advantage of this approach over traditional subjective methods.

Finally, the wealth of data points collected through facial expression capture (more than 15,000 in some single video conditions) allows for a great deal of post hoc analysis and demonstrates a marked superiority in consistency and sheer volume of actionable output for purposes of post training analysis. Additionally, the high sampling rate of these data allows for much finer granularity in terms of insights gleaned - individual engagement and affect can be assessed down to the fraction of a second, allowing finely targeted interventions to improve training engagement or to address issues of concern.

FUTURE RESEARCH

Automated assessment of affective responses through facial expression capture and analysis needs further study. In particular, a comparison to other indicators of affective responses such as self-reporting, biometrics, or validated clinical scales would help to ensure that levels of accuracy in the automated processes meet the external and construct validity benchmarks required for implementation into a viable assessment product. Facial expression data capture and analysis as a heuristic to intuit human emotion is still a relatively new technology in need of further improvement and study before the effectiveness and reliability of such techniques is clear.

Additionally, research using similar methods is currently ongoing to study the effects of engagement on knowledge transfer and retention. Though not the focus of the original two studies, evidence suggests that a point of engagement satiation may occur, where “too much” engagement may actually inhibit knowledge transfer. A trainee experiencing extreme levels of engagement might fail to retain the actual subject matter material. A better understanding of how to assess affective responses and engagement in real time could considerably reshape future approaches to “gamification” and other trending techniques to improve user engagement in digital training products.

REFERENCES

Affectiva Website (2016) [Online Webpage Content] *Technology* Retrieved June 1, 2016, from http://www.affectiva.com/technology

Bob Sottilare, B., Goldberg, B.S., & Brawner, K.(2012). Effective Learner Modeling for Computer-Based Tutoring of Cognitive and Affective Tasks. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012*, Orlando Florida

Frost D.L. & Sullivan, L.A. (2015). Analysis of the Global Military Training & Simulation Market

Mehrabian, A.; Wiener, M. (1967). Decoding of Inconsistent Communications. *Journal of Personality and Social Psychology* 6 (1): 109–114.http://www.affectiva.com/technology