# Youth Engagement during Making: Using Electrodermal Activity Data and First-Person Video to Generate Evidence-based Conjectures

Victor R. Lee

Stanford University

vree@stanford.edu

## Abstract

- Purpose

  The paper introduces and explores the use of electrodermal activity data as a tool for obtaining data about youth engagement with maker learning activities.

- Design/methodology/approach

Electrodermal activity and survey data were collected from a yearlong afterschool maker program for teens that met weekly and was hosted at a children's museum. Data from four youth who were simultaneously present for eight weeks were examined to ascertain what experiences and activities were more or less engaging for them, based on psychophysiological measure.

- Findings

Most of the focal youth appeared to show higher levels of engagement by survey measure throughout the program. However, when examined by smaller time intervals, certain activities appeared to be more engaging. Planning of maker activities was one space where engagement was higher. Completing sewing projects with minimal social interaction appeared to be less engaging. Specific activities involving common maker technologies yielded mixed levels of engagement.

- Originality

Some research is emerging that uses electrodermal activity data as a basis for generating inferences about various states while participating in maker learning activities. This paper provides a novel analysis building on some techniques established in the still emergent body of prior research in this area.

## 1.0 Introduction

Much of the enthusiasm among researchers and other proponents for maker learning activities is manifested in discourse that emphasizes how engaging those activities can be (e.g., Martin, 2015). For some, the long-term hope is that this high level of engagement leads to the development of new interests. For instance, Dougherty has suggested that engaging in making could “spark interest in STEM the arts, and learning as a whole” (p. 14, Dougherty, 2013). Others argue that engaging in making can be a means to disrupt existing power structures by repositioning learners and their experiences and goals (Greenberg et al., 2020). Making, given that emphasis, provides the opportunity to change relations youth have with larger forces and their position within communities. Still others see engagement as being important in making as it supports enactment of certain valued disciplinary practices, such as those associated with science and engineering (e.g., Simpson, et al., 2020). Based on these examples, we can see that youth engagement is invoked and appreciated by many in the maker education research community. However, what precisely is being referenced and recognized when we talk about youth engagement in making is quite variable. Sometimes, it is a proxy for participation in the visible practices of making. Sometimes, it is insinuation that the activities involved in making have a sort of ‘stickiness’ that keeps learnings interested and involved in ways that traditional forms of instruction do not.

Given its ubiquity and varied usage, a convergent definition of youth engagement as it relates to making will remain elusive. Yet despite the lack of convergence, I contend it is still worthwhile for maker educators and researchers to propose and test specific operationalizations of engagement and examine what follows analytically from those operationalizations. For example, imagine a proposal made by a researcher that we operationalize engagement by measuring how much smiling takes place during a youth maker activity. That operationalization might be the number of different youth who smile, the duration of smiles, the frequency with which smiles are elicited, or some combination. Imagine further that whatever 'measure of smiles' is selected yields analyses that show a pattern of youth smiling more (however we define 'more') when youth are sharing their accomplishments while making. From that, we can advance a new data-informed conjecture about making: providing opportunities for youth to share their accomplishments can boost youth engagement. That conjecture then informs future research and design of maker activities. Subsequent maker programs incorporate more opportunities for youth to share their accomplishments. Research and evaluation efforts that follow examine if engagement increases with the change in maker activity design. Some researchers use the 'measure of smiles' or some variant. Others use a different operationalizations. New empirical findings are generated that affirm or challenge the initial findings. Conjectures are refined and tested, and the cycle repeats.

In this hypothetical 'measure of smiles' scenario, we could also expect that some fundamental disagreements will surface. Arguments may be proffered that 'measure of smiles' does not actually measure engagement as the critics define it, or that smile measurements bias for certain conditions to be met that are not always met (some learners may smile less or some conditions prevent smiles from being made or counted). Those critiques could have legitimacy. In the best case, challenges to smile measurement will offer more than critique; they will propose specific new operationalizations as expansions or alternatives. New conjectures will emerge, and the cycle continues. We may end up discontinuing use of smile measurements or refining them. Regardless, all of these efforts – from initial operationalization and subsequent critique – represent forward progress. Our ideas about making and engagement get further refined.

I have presented this hypothetical scenario to situate this current study. The work that follows does not measure smiles, but it does propose some specific operationalizations of youth engagement during making. Some will find the operationalizations agreeable, and others will challenge them, either immediately or in the future. However, putting forward an operationalization and seeing where it leads us is a start. In the future, as dialogue ensues, the educational research community may maintain this operationalization, refine it, challenge it, or ultimately let it fade as new operationalizations and instrumentations emerge.

I have labored through this hypothetical in order to convey that entertainment of conjectures made within this paper depends on a willingness to also entertain the proposed instrumentation. The proposed instrumentation in this study is the use of a wearable device that obtains readings of psychophysiological response. My experience in presenting work like this has been that acceptance of psychophysiological response as a proxy for youth engagement is where the most questions get raised. And those questions then lead to disagreements about what anyone means by youth engagement, and then the position is put out there that it is part and parcel of the instrumentation and methods that are used to detect it. So, with all that in mind, I acknowledge and recognize that the entire premise of this paper – that engagement, or some facet of it, operationalized through measurement of psychophysiological response, yields conjectures that can be made subject to testing in future research and design of maker learning experiences – is dependent on being willing to play the game of assuming psychophysiological measures can capture something that we call engagement. As is the case for any single research study, a set of claims and evidence are put forward, and they may be challenged, refined, or refuted as more research is done in the future.

With that prefacing in mind, I can now be more direct about the contents of this paper. I present a method, based on the collection of psychophysiological data in an intact makerspace program, that others could potentially use to measure youth engagement while participating in making. The method has its advantages and disadvantages, and one aim of this paper is to be clear about what some of those are. I have done some exploratory analyses with the collected data, and those analyses are also reported here. This is intended to demonstrate some possible ways to analyze such data and also to generate conjectures for future research and design of youth engagement within makerspaces. I take as a given that conjectures have value for design and for design-based research, often comprising the underlying logics that are tested in design-based research (Sandoval, 2013). Given those commitments, this paper should be understood as is responding to the following methodological and empirical questions:

1. The methodological questions. How could psychophysiological data, specifically electrodermal activity (i.e. how well one's skin conducts electricity, abbreviated as EDA), be used to detect youth engagement in maker activities? What challenges are associated with these data and the psychophysiologically-oriented approach?

2. The empirical questions. Assuming we accept that EDA can be a proxy for engagement, what experiences in a maker program exhibit patterns of more and less engagement for youth?

A literalist might challenge the wording of the second question; they would say the question truly is what experiences elicit more and less EDA response, not which activities are more or less engaging. However, that point is analogous to saying that a survey-based study of engagement is only saying what activities yield certain kinds of survey responses, not true information about what *actually* is engaging. And if the dialogue with the literalist continues, I expect that it would all circle back to uncertainty and disagreement about what it means for something to be engaging. This touches upon some foundational issues that will be revisited in the discussion. Prior to that section, I provide a brief literature review on the prior research on engagement and the relevant background of EDA, with emphasis on earlier studies I have been part of that used similar data to this one. I then discuss the research site and data sources and the challenges encountered with the methods described here. Following that, some new analyses are presented to illustrate what conjectures could be made based on these data. After that, the requisite discussion and conclusion appear, along with a reflection of what this assumption-laden journey buys the maker education research community.

# 2.0 Literature Review

Two bodies of literature are most relevant for the current paper. Those are literature related to the construct of engagement and literature related to EDA. It is assumed the reader is familiar with the maker education literature, and if not, they are referred to a number of excellent articles, books, and book sections that include, but are not limited to, Blikstein (2018), Honey and Kanter, (2013); Peppler et al., volumes 1 and 2 (2016a,b), and Sheridan et al. (2014) along with their references and derivative works that have cited those and other comparable syntheses.

## *2.1 Engagement*

While I have spotlighted the polysemy of youth engagement, some established traditions exist in the educational research literature with respect to the study of engagement. In previous decades, the term "engagement" was used large to refer to participation and commitment as measured by school attendance and participation in school-based activities. Motivation was included in there as well. Some clarification was offered by a review from Fredricks et al., (2004), which stated that engagement could be understood as an 'investment' that had any combination or (sub)set of cognitive, behavioral, and affective dimensions. Cognitive engagement was to be about psychological exertion while completing a mental task. An example could be the cognitive work associated with solving a sudoku puzzle. Behavioral engagement referred to observables although that could be tied to different grain sizes. If one attended a class when given the option to skip it, that could be counted as behavioral engagement. On the other hand, if someone changed from a slouching to an upright posture in response during class, that too could be considered behavioral engagement. Affective engagement involved the subjective feelings one had with respect to an activity, often with intensity taken into consideration. Someone who felt happy while and because they were doing a task or horrified and unable to turn away from a scene could be said to be highly engaged affectively. Motivation was separated from engagement in that it was treated as a psychological driver toward engagement, but it was not engagement itself (Ainley, 2012).

This multidimensional approach to thinking about engagement was discussed further in Sinatra et al (2015). Sinatra et al. proposed that researchers understand research on engagement as existing on a continuum with respect to unit of analysis. There may be engagement taking place on the scale of seconds and minutes for an individual completing a research laboratory-based task, and there could be engagement that extends for several minutes or even multiple days as would be the case in a classroom engaged in rich debate that reflects productive disciplinary engagement (Engle and Conant, 2002). The continuum for describing engagement spans across what Sinatra et al. call person-oriented to person-in-context to context-oriented. The latter example, of discursively realized productive disciplinary engagement, would be context-oriented given the sociocultural orientation, as would the discourse analysis of engagement offered by Ryu and Lombardi (2015). Person-oriented would look to measures like reading times and eye-movement data as measures of engagement (e.g., Miller, 2015), whereas person-in-context would attempt to use methods such as experience sampling (e..g, Xie et al., 2019) and would examine the interaction of individual learners with their task or environment (e.g., Azevedo, 2006). The psychophysiological approach taken in this study could be considered person-oriented, but I contend that it is more a person-in-context unit of analysis. This is in large part because research takes place in a complex learning environment (rather than a researcher's lab), and the goal is to glean some sense of what kinds of activities and experiences associated with making can be engaging for youth.

### *2.2 Electrodermal activity (EDA)*

Electrodermal activity measurement has appeared in psychology and psychophysiology research as galvanic skin response or research on skin conductance levels and has typically been reserved for laboratory-based research because some of the originally used technology required physical connection of a sensor apparatus to a computer and precise environmental conditions (Dawson et al., 2007). The basic principle is that a state of increased psychological arousal yields very slight changes in skin conductivity. This is because of a sympathetic nervous system response that appears during moments that are likely to require more attention and alertness. Among other unconscious physiological responses (such as changes in pupil dilation), the sympathetic nervous system activates sweat glands in the hands and feet. This does not necessarily yield visible sweat or moisture. However, the activation of those glands leads to a change in the conductivity of the skin. When two sufficiently sensitive electrodes are in contact with a research participants' hands, the speed with which an electrical signal travels changes. An EDA response recognized and sought after in laboratory research is a characteristic rapid increase in conductivity that decays immediately after. We refer to these events as "peaks" as they look like mountain peaks when the EDA data are plotted chronologically (see Figure 1). Peaks are being interpreted here as the onset of a key arousal response. Peaks appear between 0.5 and 5 seconds after presentation of an arousing stimulus. In laboratory research, a common experimental approach would be to flash an image on a screen in front of a research participant and measure the changes in skin conductivity in the seconds after. It should be noted that not all people are equally conductive or emit clear signals. For instance, people with schizophrenia are known to not produce typical EDA signals in response to known stimuli (Gruzelier and Venables, 1972).

<INSERT FIGURE 1 AROUND HERE>

Certain conditions have, through controlled experiments, been determined as likely to produce an EDA response. Cognitively demanding tasks, such as reading a map, completing some reading comprehension activities, solving arithmetic problems, or performing a visual search are associated with increases in EDA (Nourbaksh, 2012; Setz et al., 2010; Shi et al., 2007). Cued state anxiety, such as when participants are shown unpleasant imagery or are asked to give an unplanned presentation that will be recorded and evaluated, also is associated with EDA increases (Carillo et al., 2001; Naveteur and Baque, 1987).

Since those studies have appeared, new opportunities for EDA research outside of the lab emerged when it was discovered that not only skin from the hand could produce recognizable EDA signals, but so could portions of the inner forearm near the wrist. This enabled the production of a wrist-based wearable that could store data internally and transmit it wirelessly to the cloud given modern network technologies (Poh et al., 2010). Wired systems that read from skin on one’s hand are still considered the gold standard, but research using EDA outside of lab activities became possible. In multimodal learning analytics systems, EDA has been used as one input stream for adaptive affective computing systems that detect and respond to learner affective states (Azevedo et al., 2016; D’mello et al., 2017).

### *2.2.1 Wearable-obtained EDA as a proxy of engagement*

EDA has been lauded as a promising source of data for not just arousal but also contemporary and future research on learner engagement (Azevedo, 2015). However, the commercial availability of wearable EDA wristbands, is still in its early stages of becoming established or dismissed as a tool for measuring engagement. In this section, I summarize some of the nascent work on commercial wearable EDA as indicators of engagement.

To date, only a modest number of publications report on efforts to operationalize and validate engagement with EDA. In one study, an older but now discontinued EDA wearable, the Affectiva Q Sensor, was used in a study of 51 young children interacting with an adult to determine if it could be used as a measure of engagement (Hernandez et al., 2014). The authors determined that it was indeed successful at detecting engagement when compared with researcher ratings of engagement during those interactions. Morrison et al. (2020) conducted a study used the more recent Empatica E4 device with undergraduate student participants learning about climate change and found that measured EDA was higher for items that the participating students rated themselves as being more engaged in the learning activities. A study with a small sample of students working on a physical building task inferred that E4 EDA data correlated with more active engagement, which the researchers operationalized as active contribution of new ideas to the immediate task (Furuichi and Worsley, 2018). Increased emotional engagement was identified as being modestly correlated with EDA also using the E4 device in a study of its instrumentation, although they could not infer from EDA data whether the valence was positive or negative (Borrego et al., 2019).

Some studies, however, have raised some questions about EDA's use to measure engagement. In a study of Q sensors with youth who have intellectual disabilities or were neuroatypical and participating in occupational therapy, the device was not able to distinguish between engaged behaviors and challenging behaviors, although the focus on whether the youth were engaged *in the therapeutic task* may have made it difficult to distinguish those moments of engagement with other high arousal moments (Betancourt et al., 2017). Some studies that have compared E4 devices against more traditional finger-based EDA sensors where engagement has been studied in laboratory settings have raised questions about whether wearables and finger-based EDA technologies measure the same things (Menghini et al., 2019). Other studies have some results that are mixed. Milstein & Gordon (2020) had found some times when the E4 EDA data had medium to high correlations with traditional EDA instrumentation used to infer engagement, but others where there was no correlation. On the other hand, there is also the occasional study that suggests that devices like the E4 have greater discriminatory capabilities than traditional EDA technologies (Ollander et al., 2016).

While a conclusive determination of whether wearable wrist-based EDA data from both the Q and the E4 are optimal for measuring engagement, their data have been used by a number of projects as an engagement proxy. This includes projects that have used wearables – the Q sensor in older studies and the E4 in more recent ones - to detect learner engagement in classrooms (Daily et al., 2015), augmented reality activities (Soltis et al., 2020), lectures (Di Lascio et al., 2018), and conferences (Gashi et al., 2019). There have also been some emerging lines of work examining engagement as synchrony between multiple individuals in collaborative groups (Malmberg et al., 2019). Across all of these studies, there are subtle distinctions in what each author is considering to be engagement, which as discussed earlier and articulated by Sinatra et al.’s (2015) framework, is variable depending on the particular location on a larger continuum of units of analysis. However, just as this paper opened with a ‘measure of smiles’ hypothetical, this study proceeds with the belief that we can make some research headway if we try using the assumption that EDA tells us about engagement and see what conjectures follow from it.

### *2.2.2 EDA studies of engagement during making*

Given the caveat that this is still a relatively new form of instrumentation, EDA has been used as a data source to examine engagement in making specifically. One direction has been in multimodal analysis (Worsley and Blikstein, 2018). In that study, EDA was one of multiple data sources used to analyze activities of dyads who were brought into a lab to complete a making task. The results there were promising in that the authors demonstrated feasibility of the multimodal approach coupled with machine learning and sequence mining techniques. Related work using other computational analytic techniques have appeared in other studies from that group (e.g., Worsley, 2018).

Another direction, in which this study is situated, has differed in that it has relied on the coupling of wearable EDA data and first-person camera data during making tasks that were not designed or structured by a researcher. In one study, a pair of youth participating in a multiweek makerspace program involving equipping and launching a weather balloon with sensors wore cameras that captured still images of their activity and EDA devices (Author and Colleague, 2020). The major observation was that while the pair worked together, there appeared to be distinct response profiles. Different aspects of their shared activity appeared to be more engaging for each youth. A different study (Author and colleagues, 2019) used a similar approach for examining situational engagement in two 6-week maker camps focused on youth making their own versions of pre-specified artifacts. In Author and colleagues (2019), the youth wore newer versions of EDA devices and first-person wearable cameras, but the cameras recorded video rather than still images. In addition, daily engagement surveys using a validated instrument (Science Learning Activation Lab, 2015) were also collected form participants. Key findings from that study were that there appeared to be moderate correlations between presence of EDA ‘peaks’ and cognitive and behavioral engagement. Affective engagement, as measured by the survey, did not show significant correlation with EDA. The focus in that study was on moments when a plurality of youth showed heightened EDA activity, suggesting higher levels of engagement. The most frequently recognized moments included occasions of peer socializing, interactive instruction facilitated by one of the adults leading the camp, and physical and tangible making experiences (as opposed to making done on a computer through a programming environment or media editing software).

The current study uses and focuses on the same wearables as in Author and colleagues (2019), but differs in maker program context. In this study, different approaches and directions are pursued analytically to detect experiences that registered as engaging for observed youth.

# 3.0 Data Sources

## *3.1 Participants and Makerspace*

The data for this study were collected at a makerspace housed at a science museum located in a suburb of a major city. The museum regularly held a variety of youth and family programs and had recently established a makerspace in a building located on museum grounds, but in a small building separate from the main building that housed exhibits. The makerspace was equipped with multiple 3D printers, computers, sewing machines, a laser cutter, and a large flat screen TV. At the time of this study, this was their first year in offering afterschool makerspace programs, and they had established themed yearlong programs (such as robotics) targeted for middle school students to meet weekly after school hours during the academic year. As part of a deliberate effort to appeal to and attract more young women to making, the museum decided that one of the weekdays would be themed around electronic textiles (e.g., Buechley and Qiu, 2013). Existing research has suggested that electronic textiles are better at increasing participation from women than other related maker technologies (Buechley & Hill, 2010). Youth signed up for a theme and were to attend the same weekday throughout the year. Weekly sessions were planned for roughly 1.5-2 hours.

By the second week of the program, eight adolescent youth from local middle schools (6 females, 2 males) had signed up, and all provided consent and assent. Several weeks later, two additional girls joined the program and provided consent and assent. Demographic information about the youth participants was not collected. Two paid adult staff members from the museum were present and provided consent to participate. One adult, Tim (all proper names in this paper are pseudonyms), was the inaugural coordinator of maker programs and the makerspace. Tim had previously been a full-time middle school technology education and home economics teacher who intentionally provided maker instruction to his students in a more urban location in the state until he was hired by the museum. Sheila was a visual arts student specializing in painting who assisted in the makerspace. She had no prior experience with the digital making technologies used in the makerspace.

## 3.2 Data Collection

### 3.2.1 Wearables

Each participating youth was provided with an Empatica E4 electrodermal activity wearable device and a GoPro Sessions camera worn on a chest mount system (refer to Figure 4 to see the equipment on some youth). The E4 records conductivity measurements four times per second. An hour would produce 14,400 data points for a single participant. When the participants arrived at the makerspace, a research assistant provided each youth with their assigned devices and helped them to put the devices on. On most days, a standing camera was also set up by the research assistant to obtain a third person view of activity in the space. Across the wearable camera and standing camera, over 300 video files with duration of up to 2 hours each were collected (over 3 terabytes of video data).

### 3.2.2 Survey

At the end of the day, youth completed an eight-question validated engagement survey (Science Learning Activation Lab, 2015) that asked for agreement with statements on a four point scale about the session they had just completed. Sample statements included "I felt excited" or "I focused on what we were learning". The survey also included an empty field for short responses to identify the moment that they found most engaging. The maximum possible engagement survey score was 32 and the minimum possible was 8.

### *3.3 Challenges*

In this section, I note some of the challenges that were associated with the data collection and what decisions were made for analysis as a result. This is for a couple of reasons. One is that despite the sheer volume of data that were collected, the reader may notice and wonder why only a relative fraction could be used for the analysis reported in this paper. Specifically, while the program was scheduled and recorded for 29 weeks, only 9 weeks of data met the criteria for inclusion in analysis. Additionally, this form of naturalistic, automated, and passive data collection during making with wearable devices is still quite new. Sharing some of the real-world challenges of this approach is a contribution that can be made from this work for others curious about this methodology.

#### *3.3.1 Challenge 1: Attendance was inconsistent.*

While youth were signed up to participate for the entire school year, there was a combination of participant attrition and periodic absences. The project began with 8 youth of mixed genders, which later became 10 youth. However, both male participants stopped attending after the second week. Average daily attendance was 5.48 youth per week (sd = 1.09). Because of this, I identified the youth who had the highest attendance and continued with the program for the entire school year. There were five youth, all girls, who fit this criterion. Of the five, one joined a month after the program had begun. Because that one youth started quite late, she was excluded from the rest of the analysis. For the remaining four girls, there were 19 out of 29 weeks when all four simultaneously attended. These four became the focal youth for this study.

### *3.3.2 Challenge 2: Late registrations and the consent process*

For a number of youth who signed up within just a few days of the first week of the program, consent forms could not be distributed and reviewed by the youth and their families until the first week of the program. Consent forms were signed for all participants by the second week. Because data collection could not proceed without documented informed consent, the first week was excluded from analysis as three of the focal girls' families were late in providing consent.

### *3.3.3 Challenge 3: Staffing and directional changes*

Tim, the director of the making programs at the museum, was an enthusiastic educator and very accommodating of this research project. However, about two-thirds through the year, he was recruited by some colleagues for an emergency full-time teaching position at a public school. Tim and the museum could not find a workable arrangement for him to continue with the maker program, so he announced at the beginning of the 21st week that he had resigned. On his final day, the group suspended all making activities and did other farewell activities outside of the makerspace. Sheila then became the sole staff person with the group and did not have a background in digital making technologies. The program then changed to be a series of art activities. Because of this shift, data that were collected during the 21st week and later were excluded from this analysis.

### *3.3.4 Challenge 4: Schedule restrictions for the participants*

Participation in any study is a generous use of personal time in service of research. The participants' time should be respected, and for various reasons, their time could be constrained. Some studies make strict rules for inclusion and select only participants who can meet the time commitment. The approach undertaken in this project was to conduct research in an existing makerspace program that would exist regardless of researcher involvement. For researchers familiar with projects involving youth, absences are common. For out-of-school programs, there are often constraints that exist due to transportation. The youth may have a busy parent who is picking their child up at a specific time so that they can immediately transport that child or another to a different activity with a set start time. Youth may be beholden to a public transit system schedule. These constraints meant that youth left without always completing end-of-day research protocol activities. For instance, some youth rushed out the door or skipped doing the engagement survey, which created holes in the data. To keep the survey conditions consistent, they would not be asked to complete the survey at a later time as their memory and feelings about a given day at the makerspace could change. Some youth arrived late, and that delayed the start of the entire day's maker activities. These conditions are all part of the reality of out-of-school youth programs that make comprehensive data collection challenging, even when there is one or more researcher on site to assist directly with data collection, as had been the case for this project.

#### *3.3.5 Challenge 5: Device errors and malfunction*

The Empatica E4 device, while enabling new forms of data collection, was still a relatively new commercial technology. The user interface had a single button and a single small light. Pressing the button in certain ways – such as holding the button down for different durations of time - would control all operations of the device, and the light would behave in different ways to signal that the command had been recognized. These could be, from a user-centered design standpoint, relatively meager bridges to cross the gulfs of execution and evaluation of interface design (Norman, 1989). It was also battery powered. A combination of errors in setting the devices up to record (due to confusion with the interface), youth accidentally shutting their devices off (sometimes due to the interface), and batteries failing led to 8 weeks when at least one of the four focal youth had incomplete or no EDA data. Those weeks were excluded from analysis so that the four youth could be considered consistently with one another.

### *3.3.6 Challenge 6: Different records of time*

A surprise for us that was discovered *after* the data collection had been completed was that there were multiple conflicting clocks involved in this data collection effort. The E4 has its own internal clock for which data are recorded. The GoPro camera has its own internal clock. A standing video camera also has its own clock. A clock on display or on a time-keeping wristwatch (or mobile phone) used by participants also is a separate clock. We discovered that these clocks, while all similar, were not completely in synch with one another. Incidentally, this was discovered on project computers where data were being stored (which also have their own internal clocks). Furthermore, device clocks have a tendency to drift, as they are operated internally and not communicating with the atomic clock as would happen with a mobile phone. Making these even more difficult is working with digital video files which usually do not specify time in terms of the time of day but rather specify the number of minutes and seconds of video length. This led to a dizzying and constant recalculation of what time different video events were. The lack of synchrony did not seem to exceed more than roughly 10 seconds based on our internal tests – which involved identifying landmark moments from video footage and comparing all the different recorded times - but determining precise times of certain events was virtually impossible. This is also confounded with EDA peaks appearing anywhere between 0.5 and 5.0 seconds after stimulus, making precise localization of a moment that led to EDA response impossible.

# 4.0 Analysis

## 4.1 Survey

Survey scores from the beginning of the school year until Tim's last week as makerspace facilitator were plotted. Because there was a maximum of four scores for a given day given four focal youth, no inferential test was appropriate to use. Rather, tendencies within the data were drawn from visual appraisal. All collected survey scores, regardless of whether the focal youth had acceptable EDA data, were included in this plot.

For the open-ended response to a survey question about what moment youth found most engaging, the focal youth did not consistently identify 'moments'. That is, they did not usually situate their responses to times (e.g., 'at about 5:00', 'half an hour in', 'near the end of the day'). However, it was possible to reconstruct some of the mentioned activities and identify their general time by repeatedly reviewing their first-person video footage. Many of the open-ended responses that correspond with specific days are shared in the sections below, as they provided some surprising triangulation of what was especially engaging.

## 4.2 EDA and Video

Recall that 'peaks' in EDA data are taken as indicators of an acute onset of increased engagement. With that in mind, the occurrence of peaks during certain intervals of time that corresponded to specific activities were of primary interest. A description of the algorithm used for detecting peaks and steps taken to remove potential irrelevant artifacts from the EDA data is provided in Author and colleagues (2019). To facilitate analysis, the 9 weeks of data were partitioned into 15-minute increments. This was chosen as transitions between activities initiated by Tim appeared to occur at the boundaries of those increments. Each week's session differed in duration. Video data were reviewed to exclude times that the maker program was not in session, even though some EDA data may have still been recorded.

Each 15-minute increment across the 9 eligible days were reviewed from the various camera recordings and labeled with the activity or primary maker technology that was being used. For instance, paper circuits (see Author and colleague, 2018 for a description) were the focus for four 15-minute increments (one hour equivalent). The increments were grouped together in a visualization that showed how many of the four focal youth showed EDA peak responses for each fifteen-minute increment and how many did not show peak responses. The focus of this analysis was the presence of peaks rather than the number of peaks. For increment plots that showed notable patterns, the videos were reviewed to identify what else was happening during those increments that could explain the pattern of EDA response.

# 5.0 Results

## *5.1 Survey*

Survey scores for the four focal youth (Amy, Brie, Cady, and Dawn) ranged from a minimum of 15 to a maximum of 32. In general, Amy, Brie, and Cady maintained high survey scores throughout the time period, with lower values appearing during weeks I and J. Overall. Amy, Brie, and Cady appeared to judge the entire period of time as highly engaging. Dawn showed the lowest scores and her values were noticeably lower (less than 20) during weeks F, L, M, and N. During week F, the group was making “wind wheels” which were pinwheels that were powered by connection to a phone or battery. During week L, there was a period of time observing a laser cutter in action and working on “ugly shirts” for the holidays. For week M, the group primarily planned the activities for the rest of the year together and searched the internet for ideas of projects to complete. Week N and O each involved new and distinct craft projects (making a purse out of candy wrappers and making a pillowcase).

<Insert Figure 2 Around Here>

### *5.2 EDA and Video*

Six activities were selected for this analysis. One criterion for inclusion included use of canonical maker technologies. These included E-textiles (in which the focal youth were making a decorative felt square with a battery and LED), Paper Circuits (completing and decorating paper-based guide materials demonstrating series and parallel circuits), and a Laser Cutter Demonstration. Three other activities were analyzed because they were noted as engaging in related prior research (Author and colleagues, 2019) or because they tended to have consistent EDA response patterns (most or all students showing engagement or few or no students showing engagement) across all the youth. These activities included instances of interactive instruction, when Tim would be leading a presentation or lesson but actively calling on and soliciting comments from the youth, the “franken-stuffies” activity that took place over two days and involved disassembling and then reassembling limbs on stuffed animal toys to practice sewing, and planning, when Tim would focus the youth on discussion of their future schedule and upcoming and anticipated projects and activities. The resulting plots are provided below.

The segments are labeled by week and 15-minute time segment. For instance, B-2 would refer to Week B and the second 15-minute interval of the maker session. M-4 would refer to Week M and the fourth 15-minute interval of that week's session. The bars are positioned such that the number of and specific youth who showed an EDA peak response are on the positive side (i.e., above) the x-axis, and those who did not show an EDA peak response during the time interval are shown on the negative side (i.e., below).

### 5.2.1 E-textiles

While the long-term theme for the entire year's program was e-textiles, the youth actually only spent one day working with conductive thread, battery holders, and LEDs. All other activities were intended to build foundational knowledge for a larger e-textiles project, but that larger project did not materialize in part because of Tim's departure from the program. In certain weeks, there were textile activities, such as making craft items, modifying clothing, and sewing pillowcases. However, these did not integrate circuitry or electronic components. While I would maintain that those still count as valid forms of making, the emphasis in this paper is on common forms of digital making that typically involves circuitry or digital technologies.

<Insert Figure 3 Around Here>

During the four intervals when the focal girls were working on e-textiles, the pattern of EDA peaks was consistent. Amy and Dawn consistently showed EDA peaks, while Cady and Brie showed none. In Amy's end-of-day engagement survey, her open response to the question of what moment was most interesting, she wrote, "I liked making circuts <sic> with fabric and conductive thread. It was fun to come up with a design to make. I made a christmas ‘ugly sweater’ that was about the size of my palm and I loved making it.” In Cady’s open response for most interesting moment, she wrote “when he [Tim] said that the string had metal in it and also the time in the beginning that we talked about what we were doing”. Cady’s reference to “string” was about conductive thread and Tim’s explanation for why it could conduct electricity. Note that Cady’s response referred to the beginning of the day when the materials were described and a plan was being discussed, and it did not mention the work of physically making her e-textile project.

### *5.2.2 Laser Cutter*

<Insert Figure 4 Around Here>

The laser cutter demonstration was an impromptu event where Tim was preparing something to cut for a different group’s activity. The focal youth all gathered around the laser cutter and observed what was happening in the laser cutter and then proceeded to socialize with one another and with Tim for the two 15-minute sessions. Across those sessions, every focal youth showed an EDA peak response. Recall that in Author and colleagues (2019), socializing was the most common occasion when multiple youths’ EDA data produced peaks. In her open-ended response for the most interesting moment of the day, Amy wrote “I liked watching the laser cutter.” Cady wrote “I liked when we watched the laser cutter it got kind of boring”, which is actually unclear whether she enjoyed or disliked the time watching the laser cutter. Dawn wrote “i really don't know i [sic] was not the best class time”. This was a day that Dawn rated her engagement level much lower than the other focal girls. In the video footage, the youth all looked at the laser cutter at work and glanced at it periodically, but facing one another and engaging each other in conversation near the laser cutter was a much more frequent activity while they were standing in that location.

### *5.2.3 Paper Circuits*

The paper circuits activity had four continuous intervals (an entire hour). The paper circuit activities began as following templates from the Chibitronics website (https://chibitronics.com/templates-downloads/ ) and the youth worked on serial circuits, then parallel circuits, and then switches. At the end, they had time to make their own paper circuit, with some of the youth choosing to just decorate a completed template. There appeared to be an increase over time with respect to the presence of EDA peak response from the four focal youth.

<Insert Figure 5 Around Here>

Brie showed a pattern of no EDA peaks, then peaks, then none, then peaks again across the hour. In Brie's written survey open response, she wrote, "What I thought was interesting about today was how many different ways there was <sic> to create a circuit. I learned about the 'ON' 'OFF' switch. It was very cool and I liked learning about it." Introduction of parallel circuits appeared during B-5 for Brie. Brie encountered and worked on switches during time B-8.

Dawn, who did not show an EDA response, gave the following in her written survey response: "when we where all shareing our thouths <sic> about the class". Note that she did not mention paper circuits or any sort of physical making activity. It is unclear what time she was referring to as engaging because thoughts were shared throughout the day intermixed with activities.

### 5.2.4 Interactive Instruction

<Insert Figure 6 Around Here>

Interactive instruction usually involved Tim presenting information to the youth and inviting responses from the attendees. The responses were often given as outbursts from the youth participants rather than raising hands and waiting to be called upon. The content of many of these was jocular, with some light teasing and laughter appearing in response to some questions and comments to and from one another. Overall, the four focal youth tended to have EDA peak responses. During C-1, Sheila (the visual art student and assistant facilitator) was the primary speaker, and the focal youth were all sitting reclined in their chairs. There did not seem to be a specific focus to the discussion, and only a few of the attending youth participated. During C-1, the focal youth also moved from their initial seating area to a table where they waited with some e-textile materials but had not begun any work on the e-textile project.

### 5.2.5 Franken-stuffies

The Franken-stuffies activity involved separating limbs and heads from stuffed animals and then re-attaching various parts from one stuffed animal to another. The activity took place over two weekly sessions, G and H. On G, the stuffed animals were cut during the first interval, and then the remaining 45 minutes involved girls sewing limbs back on. On day H, the beginning of the Franken-stuffies section involved dumping the detached stuffed animal parts out from a bag and finding their partially completed projects from the previous week. The remainder of the time was dedicated to sewing.

<Insert Figure 7 Around Here>

This activity stands out in the data corpus because it had mostly no peaks across all or most of the focal girls. During the time when the girls were sewing, they took their stuffed animals and sat on the couches by the television. They then spent the entire time watching YouTube videos of skits from a popular local content producer. The youth were mostly quiet and sewed and watched, as illustrated in Figure 7. On occasion, a youth would get up to go get an additional limb or material from the table and then return to her seat by the television. The open-ended responses from the youths were largely general about the day's activity. The most distinct response came from Amy: "Most of today was just tearing apart stuffed animals, and that was fun except that when we started sewing i poked my finger and that wasn't fun" (from day G).

### *5.2.6 Planning*

While unexpected when we began this study, the planning sessions appeared to fairly consistently yield EDA peaks. As context for what took place during planning and how that differed from interactive instruction, the focus in planning was on future activities beyond the current week. The planning sessions varied in content with Tim often writing the schedule on the whiteboard and thinking with the girls about when different milestones could be met for a set of large e-textiles projects.

<Insert Figure 8 Around Here>

Planning was not always mentioned in the girls' open-ended survey responses. This is not surprising as planning typically did not take more than 15 minutes and other activities took place during the day. Week M was an anomaly. It was the first day after winter break and a lot of time was dedicated to reorganizing their plan and ambitions given that they were about halfway through the year. As some additional background, Tim had intended for the girls to complete an e-textiles fashion show and to also use program funds to subsidize a trip to the Bay

Area Maker Faire to show off their projects. This ultimately did not materialize, but it was discussed during week M. He also showed videos and images of sample completed maker projects that were available online. The samples varied depending on the day's activity. For instance, on the day of paper circuits, some paper circuits projects made by others that had been publicly posted online were viewed.

On week M, planning appeared to feature in the girls' open-ended survey responses. Amy offered in all capital letters, "I LOVED TALKING ABOUT PROJECT IDEAS. IT WAS SO MUCH FUN TO BE ABLE TO GET CREATIVE ABOUT PROJECT." Dawn wrote that the most engaging moment was "when we where <sic> talking about the trip" referring to discussion about the proposed Bay Area Maker Faire trip. Cady, who did not show an EDA response during the planning, referenced in writing a later activity that was not related to planning as the most engaging moment for her during week M.

After planning was completed, Tim led the youth in discussions of ways to think about problems to solve in order to encourage some creative thinking on what sort of culminating e-textiles project to complete. Then the youths each took out a makerspace laptop and were told to search for interesting projects to further help them generate ideas. Cady's response for the most engaging moment was "I liked searching the light things <sic>", which likely referred to her time using the computers and searching on the internet for projects that could light up. This was also referenced by Brie, who offered "When I saw the cool lights on clothes.". However, Brie did have EDA peak responses during M-2 through M-4, whereas Cady did not.

# 6.0 Discussion

From the analyzed data, we can begin to make some inferences about these four girls, the instrumentation, and from those we can make some informed conjectures about how maker activities influence engagement. We are also given the opportunity to reflect on the nature of engagement and what level of generalizability we can make with a study such as this one. Finally, given a deliberate effort in this article to provide transparency regarding challenges of instrumentation and data collection, we can consider how such challenges should figure into future research of this nature.

## 6.1 Forming conjectures

One inference we can make from the data presented in this paper is that very different images can be gained from use of different methods and by focusing on different time spans. The written engagement survey was a daily indicator of engagement. It references up to two hours of experience when a number of things happened and collapses that into a single value. From review and acceptance of survey scores, three girls seemed to be fairly highly engaged overall throughout the year. We might infer that overall, the program was engaging because of those results. The only substantive deviation came from Dawn. For her, there appeared to be a decline in engagement in the latter portion of the year. The primary maker activities on those days differed in materials and tools, so her decrease in engagement does not seem to follow an easily detected pattern. That is, we cannot yet say that specific tools or specific activities are ones that we expect will be unengaging for Dawn. So, if we were to provide a takeaway just from these girls' survey data, we might infer that this program was a success at yielding higher levels of engagement. That inference is subject to critique based on sample size, data quality, and whether the instrument was adequate for measuring youth engagement in the first place.

However, the paper is not reliant exclusively on survey scores. Much of the novelty of this study was the effort to drill deeper using passively collected EDA. When we looked at electrodermal activity data across shorter time spans and aggregated those, other tendencies began to appear. From those, we could make informed conjectures. One conjecture, derived from the pattern of higher engagement during planning, is that aspirational ideation and discussion of rare, novel experiences can elicit engagement. Another conjecture, informed by analysis during the Franken-stuffies activity, is that decreasing social interaction during a maker activity – by putting something on the television that everyone quietly watches - may lead to lower levels of engagement. The youth might be said to be engaged with the videos they watched, but for makerspaces, that is not the form of engagement that is typically sought. These conjectures, built from the planning and Franken-stuffies data, are consistent with one major observation from our previous EDA engagement study (Author and Colleagues, 2019); social interaction is an important driver of engagement in youth maker programs. The sheer presence and access to specific maker technologies does not seem to determine whether a youth will be engaged. Being in the presence of or having the option to use new fabrication technologies seem to be less important than the social interactions that take place in making. Many readers would not expect that to be the case, and I am no exception. However, from this study, we have some empirical support for that position.

## 6.2 Reflecting on what "actually" is being measured

There is a larger concern that loomed over this project, and that is whether these EDA measurements and analyses actually measured the thing that we value in making and call youth engagement. What is literally being measured is how quickly electrical impulses move through the skin. Beyond that, we have to make an inferential leap as to what changes in skin conductivity indicate. Legitimate critiques could be that something other than engagement is being measured. It might have been some other construct we refer to as arousal, alertness, or stress. With that deferral of EDA responses to different constructs, the youth who were watching videos while sewing Franken-stuffies might then be considered highly engaged, but not aroused or stressed. In that case, a different operationalization of engagement is needed, and one must again return to the question of what do we wish to consider as engagement. What are acceptable proxies? Depending on the answer, the acceptance of electrodermal activity will differ. The case for EDA as being one, but not the only nor necessarily the best, has been presented through this paper.

One might reasonably ask if EDA is worth using rather than surveys or video analysis. As observed by Hernandez et al., (2014), EDA has intuitive appeal to other methods. Surveys and self-reports are subject to a recall bias where a number of factors can affect how a participant responds. For one, the participating youth may not even share the same definition of what is engaging as the researchers. Video analysis is time intensive and laborious. If a standing camera is used for observing naturally occurring activity, there is much that an analyst must infer from the video footage and the limits that come from video-based perspectives that must be considered (Hall, 2000). Some systematic video analysis approaches could be used – such as analysis of facial expressions using an established coding scheme – but at present, that too is very time intensive and subject to human judgments. There may be future artificial intelligence technologies that may help expedite that work, but it is still quite some time away from being available and its reliability and appropriate use will depend on what training corpora are used to develop those techniques. EDA should not be treated as the final authoritative data source for studying engagement – this article does not attempt to nor endorse making that claim - but EDA should be considered and explored as one of many options currently within our research toolkits.

### 6.3 Recognizing Challenges with this Type of Research

Earlier, several challenges were identified that were associated with this data collection effort. These included inconsistent youth attendance, delayed start because of late registration and consent, changes in staffing, participant schedule restrictions, device errors, and different records of time. The first four of these are endemic to educational research that seeks to work with programs in situ. Researchers are not able to, nor should be in a position to, require or control youth attendance or participation. They also should be expected to follow proper consent procedures as a matter of ethical conduct. Staffing changes were also outside of the research team’s control and in the case of this project, took everyone involved with this program by surprise. We cannot be assured how the future will unfold and must adjust based on those things. In future work, it could be wise to take staffing into consideration in that some organizations may have consistently high turnover or be early enough in their development that stable staffing arrangements have yet to be developed. However, this type of event is not uncommon, and thus problematic only in that it deviates from what would have been our ideal.

The challenges where researchers can insist on more changes are the latter two, related to the specific EDA instruments. We should expect that device designers and manufacturers will have thought about user experience and use cases in some detail. By even publishing this study and publicizing its challenges, we can provide some documentation for device designers that will help inform their future work. Human error is to be expected with all manner of technologies and processes and should be considered as part of the design process (Norman, 1988). Considering we are still in early iterations of wearable EDA technology, we are tolerant of these frustrations and hopeful that future device releases will better address these problems. Any near-term research that uses the technologies used here would be well advised to plan for ways to mitigate some of the device-based challenges. Future research should also be very cautious with respect to attributing increased arousal to a specific moment and would be well advised to keep attributions at the level of activities.

## 6.4 Generalization

If, with warts and all, we do deem that this methodological program of using EDA to detect engagement is acceptable, there is still room for other concerns to be voiced. For example, the size of the analyzed sample could be critiqued. Despite having up to 9 youth and 29 weeks of data, this paper only honed in on 4 youth over 8 weeks. It is hard to generalize to the entire population that was part of this maker program, let alone to other maker programs and makerspaces beyond this one given that number of youth and number of sessions. However, that kind of broader population inference is a degree of generalization that I do not wholeheartedly endorse from this study. The type of generalization that I do support and see as being made is in the form of evidence-based conjecture. Should those conjectures be integrated into future research and design, then the findings of this study are being generalized. That is the sort of generalization that is often associated with case study methodology (Flyvbjerg, 2006). We look to such research to provide us with new and plausible ideas that help us further our ability to understand the world. It would have been desirable to have more youth and more days to analyze, but the reality of the data context made it such that those that met some conservative criteria of who could be studied were the ones studied.

Still, given the limitations, this effort was intended to move the work of understanding engagement in makerspaces forward. There are a number of ways such work could be reconfigured in service of fostering greater levels of engagement. For instance, the EDA data could be returned to the youth immediately for them to know how their bodies were responding and to use that information to shape their choice of future activities. Increased EDA levels across youth could be monitored and send an alert to a facilitator when engagement, operationalized through skin conductance, is low or dropping. As of now, the initial research on wearable EDA detection for measuring engagement is still quite new. Future work and uptake will determine whether or not this approach is seen as having sufficient purchase for more progress to be made in our understanding of when and how maker activities are engaging for youth.

## Acknowledgments

The author acknowledges the valuable contributions of <redacted for review>. This work was supported in part by funding from the National Science Foundation under Grant

No.<redacted for review>. The opinions expressed herein are those of the authors and do not necessarily reflect those of the National Science Foundation.

## References

AUTHOR & COLLEAGUE. (2018)

AUTHOR & COLLEAGUE. (2020).

AUTHOR & COLLEAGUES. (2019)

AINLEY, M. 2012. Students' Interest and Engagement in Classroom Activities. *In:* CHRISTENSON, S. L., RESCHLY, A. L. & WYLIE, C. (eds.) *Handbook of Research on Student Engagement.* Boston, MA: Springer US.

AZEVEDO, F. 2006. Personal Excursions: Investigating the Dynamics of Student Engagement. *International Journal of Computers for Mathematical Learning,* 11**,** 57-98.

AZEVEDO, R. 2015. Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist,* 50**,** 84-94.

AZEVEDO, R., TAUB, M., MUDRICK, N., FARNSWORTH, J. & MARTIN, S. A. 2016. Interdisciplinary Research Methods Used to Investigate Emotions with Advanced Learning Technologies. *In:* ZEMBYLAS, M. & SCHUTZ, P. A. (eds.) *Methodological Advances in Research on Emotion and Education.* Cham: Springer International Publishing.

BETANCOURT, M. A., DETHORNE, L. S., KARAHALIOS, K. & KIM, J. G. 2017. Skin Conductance as an *In Situ* Marker for Emotional Arousal in Children with

Neurodevelopmental Communication Impairments: Methodological Considerations and Clinical Implications. *ACM Trans. Access. Comput.,* 9**,** Article 8.

BLIKSTEIN, P. 2018. Maker Movement in Education: History and Prospects. *In:* DE VRIES, M. J. (ed.) *Handbook of Technology Education.* Cham: Springer International Publishing.

BORREGO, A., LATORRE, J., ALCAÑIZ, M. & LLORENS, R. Reliability of the Empatica E4 wristband to measure electrodermal activity to emotional stimuli. 2019 International Conference on Virtual Rehabilitation (ICVR), 21-24 July 2019 2019. 1-2.

BUECHLEY, L. & HILL, B. M. 2010. LilyPad in the wild: how hardware's long tail is supporting new engineering and design communities. *Proceedings of the 8th ACM Conference on Designing Interactive Systems.* Aarhus, Denmark: Association for Computing Machinery.

BUECHLEY, L. & QIU, K. 2013. *Sew Electric,* Cambridge, MA, HLT Press.

CARRILLO, E., MOYA-ALBIOL, L., GONZÁLEZ-BONO, E., SALVADOR, A., RICARTE, J. & GÓMEZ-AMOR, J. 2001. Gender differences in cardiovascular and electrodermal responses to public speaking task: the role of anxiety and mood states. *International Journal of Psychophysiology,* 42**,** 253-264.

DAILY, S. B., JAMES, M. T., ROY, T. & DARNELL, S. S. 2015. Engageme: Designing A Visualization Tool Utilizing Physiological Feedback To Support Instruction. *Technology, Instruction, Cognition and Learning,* 10**,** 107-126.

DAWSON, M. E., SCHELL, A. M. & FILION, D. L. 2007. The electrodermal system. *In:* CACIOPPO, J. T., TASSINARY, L. G. & BERNTSON, G. G. (eds.) *Handbook of Psychophysiology.* Cambridge, MA: Cambridge University Press.

DI LASCIO, E., GASHI, S. & SANTINI, S. 2018. Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.,* 2**,** 1-21.

D'MELLO, S., DIETERLE, E. & DUCKWORTH, A. 2017. Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational Psychologist,* 52**,** 104-123.

DOUGHERTY, D. 2013. The maker mindset. *In:* HONEY, M. & KANTER, D. (eds.) *Design, Make, Play: Growing the Next Generation of STEM Innovators.* New York, NY: Taylor & Francis.

ENGLE, R. A. & CONANT, F. R. 2002. Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition & Instruction,* 20**,** 399-483.

FURUICHI, K. & WORSLEY, M. 2018. Using Physiological Responses To Capture Unique Idea Creation In Team Collaborations. *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing.* Jersey City, NJ, USA: Association for Computing Machinery.

FLYVBJERG, B. 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry,* 12**,** 219-245.

FREDRICKS, J. A., BLUMENFELD, P. C. & PARIS, A. H. 2004. School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research,* 74**,** 59-109.

GASHI, S., LASCIO, E. D. & SANTINI, S. 2019. Using Unobtrusive Wearable Sensors to Measure the Physiological Synchrony Between Presenters and Audience Members. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.,* 3**,** Article 13.

GREENBERG, D., CALABRESE BARTON, A., TAN, E. & ARCHER, L. 2020. Redefining entrepreneurialism in the maker movement: A critical youth approach. *Journal of the Learning Sciences*, 1-40.

GRUZELIER, J. H. & VENABLES, P. H. 1972. Skin conductance orienting activity in a heterogeneous sample of schizophrenics: Possible evidence of limbic dysfunction. *The Journal of Nervous and Mental Disease,* 155, 277-287.

HALL, R. 2000. Videorecording as theory. *In:* LESH, R. & LEHRER, R. (eds.) *Handbook of Reserach Design in Mathematics and Science Education.* Hillsdale, NJ: Lawrence Erlbaum.

HERNANDEZ, J., RIOBO, I., ROZGA, A., ABOWD, G. D. & PICARD, R. W. 2014. Using electrodermal activity to recognize ease of engagement in children during social interactions. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* Seattle, Washington: Association for Computing Machinery.

HONEY, M. & KANTER, D. E. 2013. *Design, Make, Play: Growing the Next Generation of STEM Innovators,* New York, NY, Routledge.

MALMBERG, J., JÄRVELÄ, S., HOLAPPA, J., HAATAJA, E., HUANG, X. & SIIPO, A. 2019. Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior,* 96, 235-245.

MARTIN, L. 2015. The promise of the Maker Movement for education. *Journal of Pre-College Engineering Education Research (J-PEER),* 5, 4.

MENGHINI, L., GIANFRANCHI, E., CELLINI, N., PATRON, E., TAGLIABUE, M. & SARLO, M. 2019. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology,* 56**,** e13441.

MILLER, B. W. 2015. Using Reading Times and Eye-Movements to Measure Cognitive Engagement. *Educational Psychologist,* 50**,** 31-42.

MILSTEIN, N. & GORDON, I. 2020. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in behavioral neuroscience,* 14**,** 148-148.

MORRISON, A. L., ROZAK, S., GOLD, A. U. & KAY, J. E. 2020. Quantifying student engagement in learning about climate change using galvanic hand sensors in a controlled educational setting. *Climatic Change,* 159**,** 17-36.

NAVETEUR, J. & FREIXA I BAQUE, E. 1987. Individual differences in electrodermal activity as a function of subjects' anxiety. *Personality and Individual Differences,* 8**,** 615-626.

NORMAN, D. A. 1988. *The psychology of everyday things,* New York, Basic Books.

NORMAN, D. A. 1989. Cognitive Engineering. *In:* NORMAN, D. A. & DRAPER, S. W. (eds.) *User Centered System Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

NOURBAKHSH, N., WANG, Y., CHEN, F. & CALVO, R. A. 2012. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. *Proceedings of the 24th Australian Computer-Human Interaction Conference.* Melbourne, Australia: ACM.

OLLANDER, S., GODIN, C., CAMPAGNE, A. & CHARBONNIER, S. A comparison of wearable and stationary sensors for stress detection. 2016 IEEE International

Conference on Systems, Man, and Cybernetics (SMC), 9-12 Oct. 2016 2016. 004362-004366.

PEPPLER, K., HALVERSON, E. & KAFAI, Y. B. 2016. *Makeology: Makerspaces as learning environments (Volume 1),* New York, Routledge.

PEPPLER, K., HALVERSON, E. R. & KAFAI, Y. B. 2016. *Makeology: Makers as Learners (Volume 2),* New York, Routledge.

POH, M.-Z., SWENSON, N. C. & PICARD, R. W. 2010. A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Biomedical Engineering,* 57**,** 1243-1252.

RYU, S. & LOMBARDI, D. 2015. Coding Classroom Interactions for Collective and Individual Engagement. *Educational Psychologist,* 50**,** 70-83.

SANDOVAL, W. 2013. Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences,* 23**,** 18-36.

SCIENCE LEARNING ACTIVATION LAB 2015. Measures Technical Brief: Engagement in Science Learning Activities version 3.1 August 2015.

SETZ, C., ARNRICH, B., SCHUMM, J., MARCA, R. L., TRSTER, G. & EHLERT, U. 2010. Discriminating stress from cognitive load using a wearable EDA device. *Trans. Info. Tech. Biomed.,* 14**,** 410-417.

SHERIDAN, K., HALVERSON, E. R., LITTS, B., BRAHMS, L., JACOBS-PRIEBE, L. & OWENS, T. 2014. Learning in the Making: A Comparative Case Study of Three Makerspaces. *Harvard Educational Review,* 84**,** 505-531.

SHI, Y., RUIZ, N., TAIB, R., CHOI, E. & CHEN, F. 2007. Galvanic skin response (GSR) as an index of cognitive load. *CHI '07 Extended Abstracts on Human Factors in Computing Systems.* San Jose, CA, USA: ACM.

SIMPSON, A., BURRIS, A. & MALTESE, A. 2020. Youth's Engagement as Scientists and Engineers in an Afterschool Making and Tinkering Program. *Research in Science Education,* 50**,** 1-22.

SINATRA, G. M., HEDDY, B. C. & LOMBARDI, D. 2015. The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist,* 50**,** 1-13.

SOLTIS, N. A., MCNEAL, K. S., ATKINS, R. M. & MAUDLIN, L. C. 2020. A novel approach to measuring student engagement while using an augmented reality sandbox. *Journal of Geography in Higher Education,* 44**,** 512-531.

XIE, K., HEDDY, B. C. & VONGKULLUKSN, V. W. 2019. Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology,* 59**,** 101788.

WORSLEY, M. 2018. (Dis)engagement matters: identifying efficacious learning practices with multimodal learning analytics. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge.* Sydney, New South Wales, Australia: Association for Computing Machinery.

WORSLEY, M. & BLIKSTEIN, P. 2018. A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education,* 28**,** 385-419.

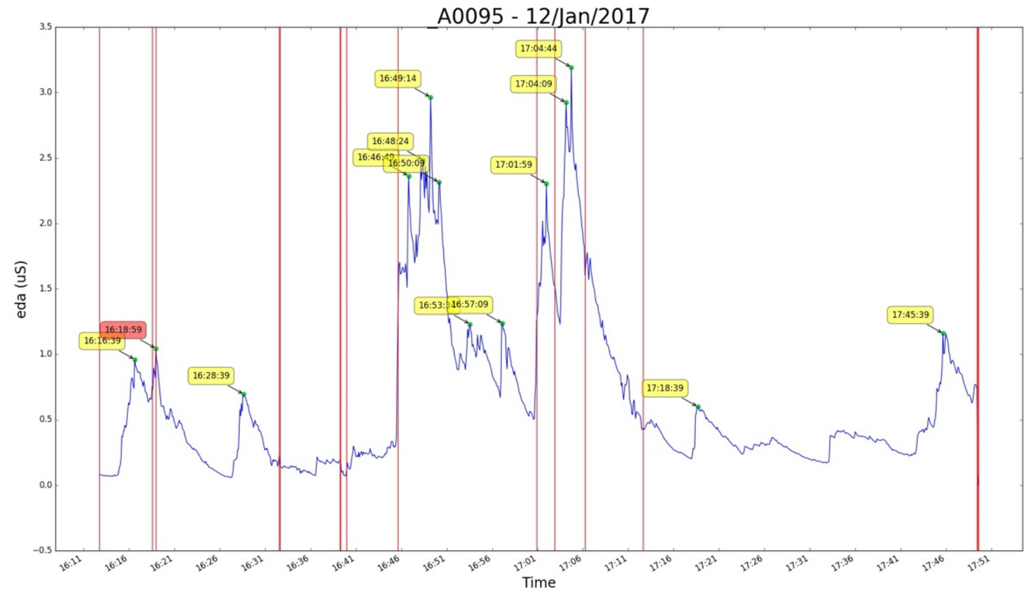


Figure 1. Approximately 2 hours of EDA data from a youth with algorithmically-identified peaks and their time of occurrence highlighted. For details on the algorithm, refer to Author & Colleauges (2019). Vertical lines indicate possible artifacts in the data identified algorithmically which were excluded from analysis if any identified peaks coincided with them.

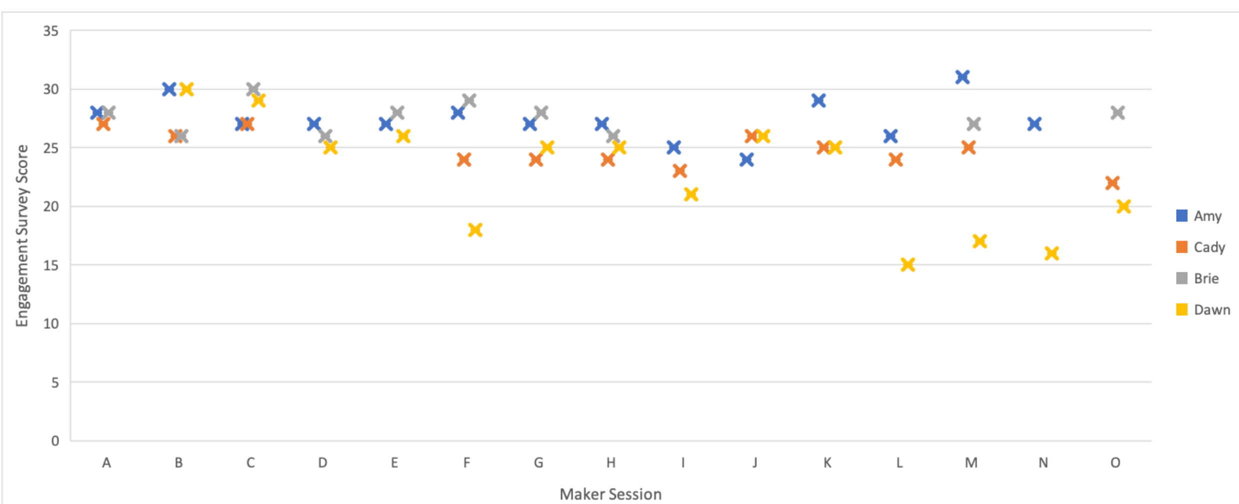


Figure 2. A plot of youth engagement survey scores across 15 sessions (range is 8 to 32 with higher meaning more engagement).


Figure 3. EDA Engagement plot for e-textiles sessions where youth made felt patches with lights (left). A still image from Dawn’s worn camera while working on felt e-textile activity during time segment C-5 (right).

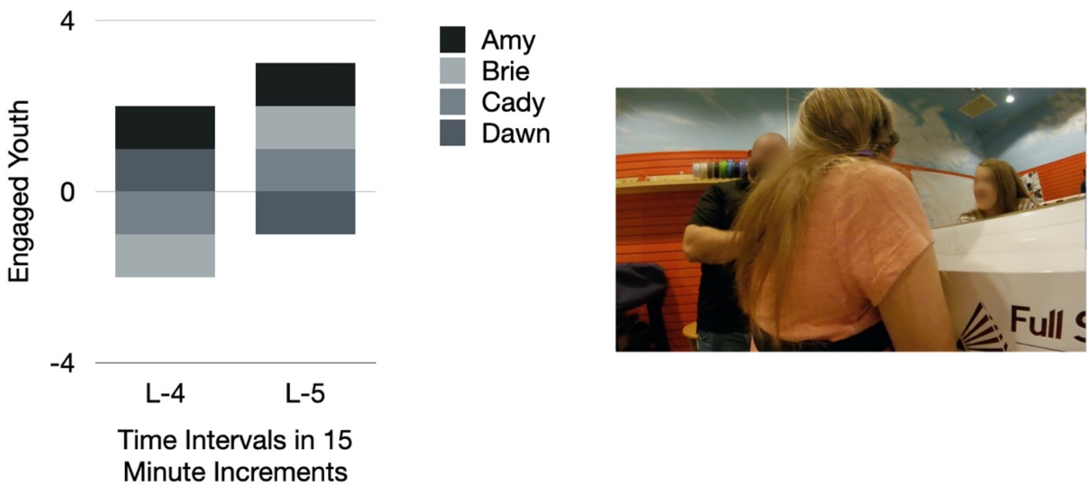


Figure 4. EDA Engagement plot for laser cutter observation sessions where youth gathered around a laser cutter in use and talked (left). A still image from Cady's worn camera as she is waiting by the laser cutter while it is in use and talking with others during L-5 (right).

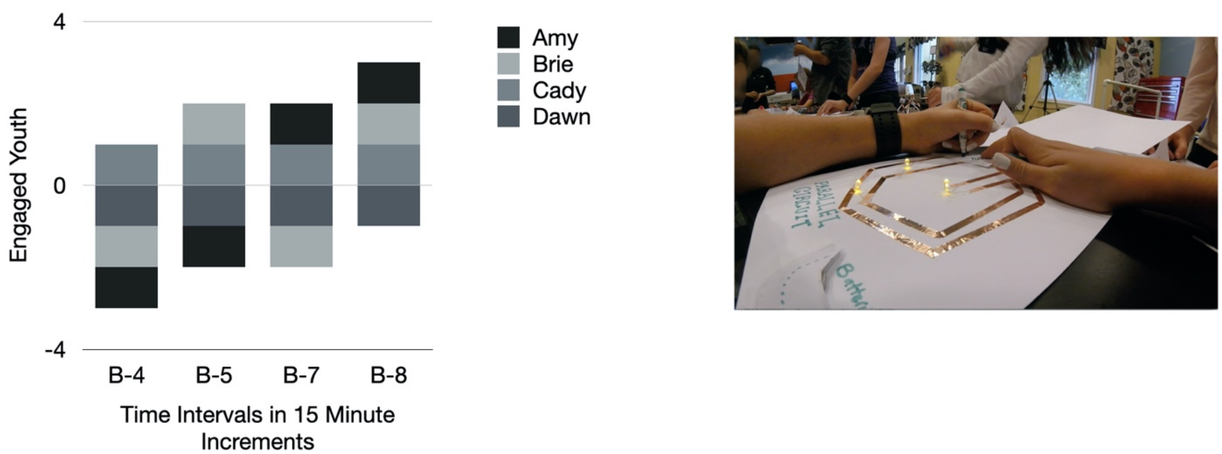


Figure 5. EDA Engagement plot for paper circuit sessions where youth completed paper circuit templates and then decorated their own paper circuit (left). A still image from Brie's worn camera as she is working on paper circuit activity for parallel circuits during time B-8 (right).

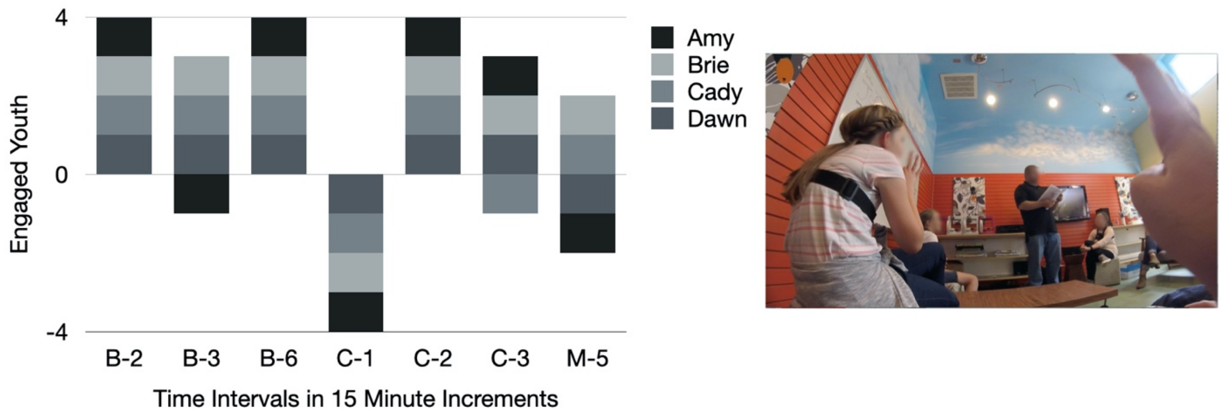


Figure 6. EDA Engagement plot for interactive instruction sessions where Tim and youth went over new content or tasks (left). A still image from Amy’s camera showing her gesturing while the facilitator talks and reviews a handout for the group during B-3 (right).

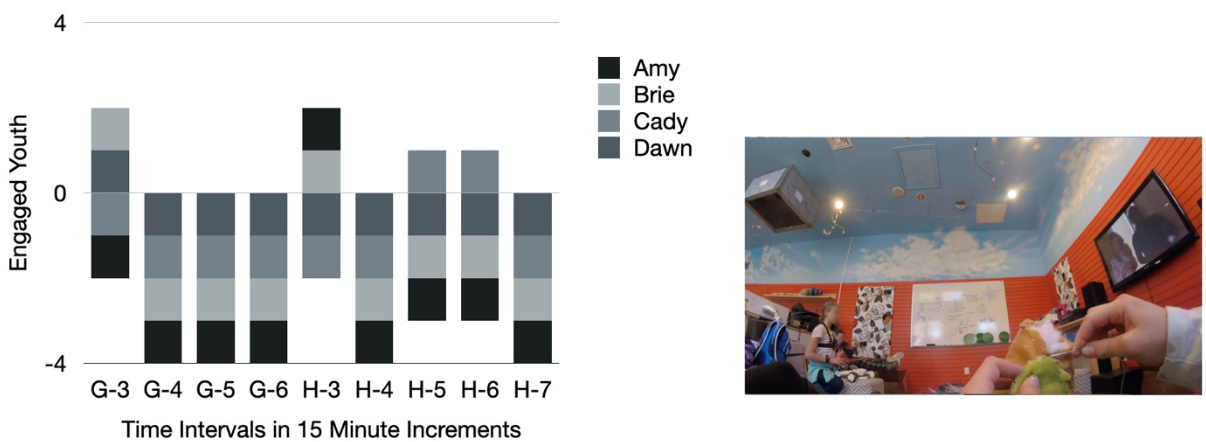


Figure 7. EDA Engagement plot for Franken-stuffies sessions where youth disassembled and reassembled stuffed animals to practice sewing (left). A still image from Cady’s worn camera when working on her Franken-Stuffie while watching a video during C-5 (right).

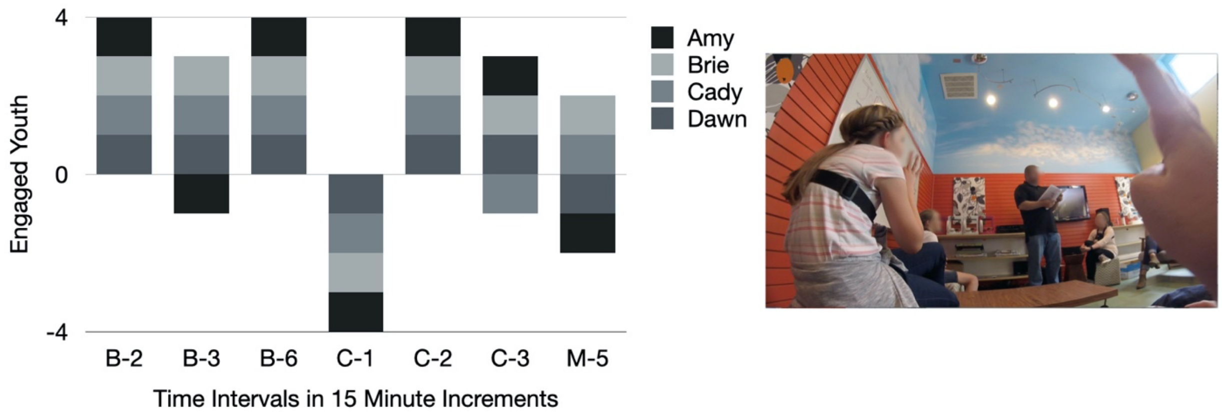


Figure 8. EDA Engagement plot for interactive instruction sessions where Tim and youth went over new content or tasks (left). A still image from Amy's camera showing her gesturing while the facilitator talks and reviews a handout for the group during B-3 (right).