Contemporary Educational Psychology 59 (2019) 101789

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

A wearables-based approach to detect and identify momentary engagement in afterschool Makerspace programs

Victor R. Lee[a,*], Liam Fischback[b], Ryan Cain[c]

[a] Stanford University, United States
[b] Utah State University, United States
[c] Weber State University, United States

ARTICLE INFO

Keywords:
Electrodermal activity
Wearables
Engagement
Makerspaces
Afterschool programs

ABSTRACT

Educational activities and programs associated with the Maker movement, which emphasizes creation of physical and digital artifacts as part of the learning experience, are presumed to be highly engaging for youth. However, there has been limited research examining what features of Maker learning activities are associated with youth engagement. We describe a research approach using wearable electrodermal activity sensors and wearable cameras to obtain data from two afterschool programs at a community Makerspace for adolescent girls (N = 12, 13). Using data obtained from these two sources along with daily survey data, we compare what is revealed from these different data sources. We observe a moderate correlation between electrodermal activity and engagement from survey responses. We also observe that activities emphasizing personal expression elicited engagement from many youth. From the EDA data and first-person video, we also identify 23 moments when groups of participants had several concurrent EDA responses suggesting high levels momentary engagement. Key features associated with those moments were opportunities for peer socialization, interactive instructional discourse, and physical making activities when objects were being assembled and manipulated.

1. Introduction

Throughout the last decade, the “Maker Movement” has generated enthusiasm from educators, funders, and researchers who have viewed the associated practices of physical and digital artifact production as offering a critical opportunity to both reshape and reimagine both formal and informal learning experiences. Much of that enthusiasm is due to very optimistic discourse around Making, which has asserted that Maker-oriented learning activities are highly engaging for students (Hsu, Baldwin, & Ching, 2017; Martin, 2015), and rich with high-interest moments and situations that will lead to the subsequent development of interests in STEM content and practices (e.g., Dougherty, 2013). However, we currently have modest empirical support of such claims.

The underlying theory of how engagement is related to Making is weakly specified. By their nature, some Maker activities “support youth autonomy and control of their endeavors” and thus will “support engagement and persistence” (Martin, 2015 p. 37). However, we do not know with any specificity which Maker activities provide for this. To illustrate, consider that Bevan (2017) has identified three different forms of educative Maker activities (assembly, creative construction, and tinkering) that have appeared in learning settings that place different demands on learners and facilitators. We may reasonably suspect that each of those may promote different levels of engagement with different populations and different circumstances. Moreover, we do not know if it is strictly autonomy-promoting activities that promote engagement. It could be that the canonical activities of Making – such as using fabrication tools and creating a new artifact - are some of the many possible contributors to engagement. Other factors could include the change from standard educational formats of classroom based instruction, the opportunities to integrate more peer socializing and collaboration during a learning activity, or some other factors that have yet to be identified.

Additionally, educational psychology has a long history of identifying variables that influence how people will engage with learning material and learning activities that can include topical and situational interest, confidence in relevant prior knowledge, self-efficacy, and self-regulation (e.g., Alexander, Jetton, & Kulikowich, 1995; Bernacki & Walkington, 2018; Cordova, Sinatra, Jones, Taasoobshirazi, & Lombardi, 2014; Linnenbrink-Garcia, Patall, & Messersmith, 2013; Moos & Azevedo, 2008). These are not addressed in Maker education literatures. As Maker education research matures, there is an

* Corresponding author at: Graduate School of Education, Stanford University, 485 Lasuen Mall, Stanford, CA 94305, United States.
E-mail address: vrlee@stanford.edu (V.R. Lee).

https://doi.org/10.1016/j.cedpsych.2019.101789

Available online 08 July 2019
0361-476X/

opportunity to connect to other empirically-established bodies of research and thus refine understandings about how engagement is established and maintained.

Yet, this has not yet been pursued. Part of the reason why a look at what moments elicit engagement has not been pursued can be attributed to the relative recency of systematic research on Maker education experiences (e.g. Chu, Schlegel, Quek, Christy, & Chen, 2017). Much of the necessary preceding research in Maker education has involved proposing different models or identifying some of the learning gains that are possible. We may only now be at a point in time when questions about moments of engagement are sensible to begin answering. In addition, the lack of work on identifying moments that support engagement are attributable to the inherent difficulty associated with detecting and identifying triggered moments of youth engagement in situ.

Recognizing this, Sinatra, Heddy, and Lombardi (2015) proposed as part of a special issue on the topic on youth engagement in science education that in order to make progress in studies of learner engagement, researchers should consider engagement as existing on a continuum that ranges from person-oriented to person-in-context to context-oriented. The specific positioning an analyst takes with respect to engagement on that continuum invites different techniques for documentation and measurement, with in situ research demanding newer techniques and approaches than the lab-based person-oriented approaches that are most prominent in educational psychology research. As a demonstration of how one could do person-in-context work to study engagement, Renninger and Bachrach (2015) argued that a move beyond self-report measures and toward more observational methods to identify what features of a learning environment trigger moments of engagement would be especially profitable. The systematic analysis of observational records, whether completed on notes or video footage, could enable researchers to capture triggers in situ and complement other established research techniques. However, they note that while such work yields valuable findings, that approach is costly in terms of time and effort.

The current paper proposes a new approach that uses an emerging sub-genre of mobile devices (Lee, 2017) as tools to help document "person-in-context" engagement in afterschool Maker learning activities. The sub-genre of mobile device that we use are wearables, and specifically electrodermal activity sensors and point-of-view cameras. The work is exploratory in nature in that this combination of technologies is a new endeavor for maker education research, and to some extent, research on engagement. Below, we describe how research on engagement has been addressed previously and how our use of electrodermal activity data is intended to provide means for us to identify what moments triggered increased youth engagement.

1.1. Engagement

As it appears in the educational research literature, engagement is a term that has been popularly used but only loosely specified. It had at one point been used primarily to refer to commitment to participation as measured by school attendance and motivation to be involved in academic activities but has since been better specified as a term that can refer to any or all of the behavioral, cognitive, and/or affective dimensions during a learning activity (Fredricks, Blumenfeld, & Paris, 2004). A common thread across these features has been an intuitive sense of an individual's "investment" in a particular activity. Behavioral engagement tended to examine participation, whether that was measured by attendance or more immediate indicators such as body position and furrowed brows; cognitive engagement referred to a psychological exertion of effort to understand ideas or complete a mental task; and affective referred to having some more intense feelings toward an experience. Summary discussions of engagement have noted contact with motivation theories, but engagement has still been treated separately; motivation has been referred to as referring to the underlying psychological processes that yields engagement (Ainley, 2012). One may have strong self-efficacy beliefs that contribute to them being more engaged in a problem solving task, but that engagement would not be considered the motivational process itself.

The construct of "interest" has also been invoked as being related to engagement, although it is not an essential component (Renninger & Hidi, 2016). As a construct, interest refers to "both the psychological state when engaging with some content and also the cognitive and affective motivational predisposition to re-engage with that content over time" (ibid., p 8). This definition of interest implies that interest has a more positive valence, but that which is engaging can still attract attention and behavioral changes, more cognitive effort, and negative affect. For example, seeing a gory injury on an internet video may lead to immediate engagement but may simultaneously elicit negative feelings and thus not incline someone to voluntarily re-engage with that video. That is, we may not expect that person to re-watch it or become especially interested in gory accidents unless there was some positive affective state that was associated with it as well. In the four-phase model of interest development, brief moments of contextually-cued engagement would overlap substantially with the first phase of interest development, dubbed situational interest (Hidi & Renninger, 2006). For instance, a child may see movie trailer about dinosaurs on the television and that may lead to future and continued engagements with dinosaurs as the interest develops. She may go see the movie, read books about dinosaurs from the library, and lobby for her family to go to the dinosaur exhibit at a local museum. However, that same moment may have also been encountered by another child who was situationally interested in the movie trailer but may not go on to see the film nor be attuned to getting dinosaur books at the library. The film trailer was interesting at the time but did not develop into a longer term interest. Still another child may have seen the movie trailer and been attentive but also terrified. That last child may deliberately opt not to learn more about dinosaurs even if given the opportunity. In all scenarios, we would consider the movie trailer to have been momentarily engaging. In that last scenario, we would consider that moment to have been momentarily engaging but may not consider it to be a situational interest.

What is common in our discussion of how motivation and interest research has treated engagement is that behavioral, cognitive, and affective dimensions can all be represented and the time scale of engaging experiences is one of seconds to minutes. That represents a particular level of granularity, and it represents a specific placement on the continuum of how engagement is conceptualized as is recommended by Sinatra et al. (2015). An experience like an extended multiweek science unit may support productive disciplinary engagement (Engle & Conant, 2002) and legitimately be called a form of engagement, but that is a different grain size and overarching perspective than the one that we are orienting towards here. That would be considered more representative of the context-oriented end of engagement. On the context-oriented end, there tend to be sociocultural analyses of the learning environment and observations of behavior, discourse interactions, and social relations in situ (e.g., Ryu & Lombardi, 2015). Critical ethnographic research on makerspace use has documented how long term activities in makerspaces can be engaging with respect to what youths value and what represents local concerns in their communities (Barton & Tan, 2018). On the other end of the continuum is person-centered perspectives on engagement, where research tends to look at an individual in a laboratory setting given a specific stimulus and measuring engagement through techniques such as self-reports, response time, or eye movements (e.g. Miller, 2015). The current paper is situated in between those two extremes, or what Sinatra et al. (2015) refer to as "person-in-context". However, our methods exhibit deliberate orientation toward some multimodal person-oriented approaches such as facial expressions and psychophysiological responses. We examined the psychophysiological response that can be detected as changes in electrodermal activity in conjunction with facial expressions to further determine affective states of the person-in-context. Multimodal analysis of converging data provides new and/or confirmatory insights that single data streams cannot always provide (Grafsgaard et al., 2014; Sawyer, Smith, Rowe, Azevedo, & Lester, 2017).

1.2. Electrodermal activity

Azevedo (2015) has noted that electrodermal activity (EDA) measurements are an especially promising approach for looking at engagement that takes place on short time scales. Similarly, advocacy for metrics like EDA have appeared elsewhere especially as they could be a data stream for automated and adaptive learning systems (D'Mello, Dieterle, & Duckworth, 2017) and in multimodal learning analytics (Azevedo, Taub, Mudrick, Farnsworth, & Martin, 2016). Until recently, EDA research has been affiliated more with the subfield of psychophysiology where it had been also referred to as galvanic skin response and been subject to laboratory studies, as precise conditions were needed for instrumentation purposes (Dawson, Schell, & Filion, 2007).

What is currently known about EDA is that changes in EDA are associated with heightened levels of arousal. The mechanism for this is tied to the sympathetic nervous system, which in response to situations that require or otherwise invite more attention, immediately activates sweat glands near a person's hands and feet to prepare the body for action (Matsumoto, Walker, Walker, & Hughes, 1990). The influx of conductive liquid from the sweat glands yields a measurable change in conductivity even without any sweat reaching the skin's surface, nor needing for there to be enough sweat to produce a visible droplet. This typically involves an immediate positive increase in conductivity in the range of 1–3 microSiemens (mS) that then rapidly decays. The onset of this reaction is between 0.5 and 5 s after introduction of stimulus. Not all individuals are equally conductive, and there are some known populations, such as individuals who have been diagnosed with schizophrenia, who do not produce typical EDA readings or responses (Gruzelier & Venables, 1972).

Cued state anxiety is one known cause of increases in EDA. For instance, when undergraduate students were shown disturbing images such as disfigured babies, bloody faces, or insects, they showed increased EDA compared to when they were shown neutral imagery such as landscapes (Naveteur & Baque, 1987). However, individuals who were measured as having trait anxiety tended to show smaller increases in EDA when shown the same disturbing stimuli. Anxiety and associated increased EDA response has also been observed in tasks where participants are asked to give a lengthy speech on an unfamiliar topic that they are told will be recorded and evaluated for a grade (Carrillo et al., 2001), a situation that is expected to provoke anxiety and stress.

High cognitive load also has been associated with increased EDA. Tasks such as map reading, arithmetic problem solving, reading comprehension, and visual search and classification tasks have been used to create increasing cognitive task demands and have shown higher levels of EDA among research participants (Nourbakhsh, Wang, Chen, & Calvo, 2012; Setz et al., 2010; Shi, Ruiz, Taib, Choi, & Chen, 2007). That suggests EDA can be an indicator of cognitive engagement. With more complex stimuli that have numerous activities appearing, EDA has been used effectively to monitor increased engagement. For example, audience response to performance arts videos have been measured with EDA (Latulipe, Carroll, & Lottridge, 2011). In classroom based research, student EDA conductivity levels has been considered a reasonable proxy for engagement (Daily, James, Roy, & Darnell, 2015; Di Lascio, Gashi, & Santini, 2018). In these operationalizations of engagement, there is a recognition that the sensors used detect arousal and are thought to be a proxy for attention, cognitive effort, and some sort of affective response although the precise valence is not known. Given the above discussion of engagement, we deem EDA as at least worthy of further exploration for detecting engagement.

Historically, and in several of the cited studies, EDA has been obtained following a person-centered research in a laboratory setting with participants wearing electrodes wired to a computer. For our current work, we have been using new wearable electrodermal activity (EDA) wristbands that allow for data to be collected in settings beyond the laboratory (Poh, Swenson, & Picard, 2010). Initial evaluation by Poh et al. suggests concurrent validity with other accepted EDA instrumentation. The wearable device is advantageous for the current study because it not only frees the participant from the laboratory, but it also provides full use of one's hands. Given our focus on afterschool Makerspace experiences, where we expected youth to be freely moving about and actively using their hands to complete their work, this was desirable.

1.3. Wearable devices as a form of mobile technology

In recognition of the larger special issue theme that examines mobile technology in education, it is important that we situate this study with respect to mobile technology. Wearables have been considered to be an emerging sub-genre of mobile devices (Lee, 2017). Recent instantiations of wearables designed for learning settings have tended thus far to emphasize their ubiquitous presence and hands-free usage. Wearables have also been identified as providing particular forms of support for learning, including supporting learners in personal expression, to integrate digital information into social interactions, to enable educative role play, and for just-in-time notifications (Lee & Shapiro, 2019). More recently, they have been recognized as tools that can capture records of bodily activity (ibid.). In education research, these have been used for explicitly pedagogical purposes. The data obtained from a wearable serves as an object to examine and discuss in a learning activity (Kang et al., 2016; Lee, Drake, & Thayne, 2016; Lyons, 2015). For our research purposes, they are not being used for pedagogy. Rather, the wearable platform is used to produce bodily records strictly for the purpose of academic research and analyst inspection.

Some of the theoretical advantages of wearables have been touched upon by Umphress and Sherin (2015) in their discussion of wearable cameras as research tools (e.g., Sherin, Russ, Sherin, & Colestock, 2008). Specifically, wearable cameras afford *perspective coherence* with the research participant. The researcher is able to obtain records that can aid in the reconstruction of what participants had personally experienced given the perspective afforded by a wearable device. Still, there are some inevitable limitations in that while the perspective coheres more with what a participant experienced, the perspective of the wearable is still a forced perspective given where the device was placed on the body and what it is instrumented to record. For instance, a wrist-born wearable step tracker can do a respectable job of tracking steps over the course of a day (Diaz et al., 2015), but the number that is ultimately reported by the wearable device is an approximation for the actual steps taken since the state of the art does not allow all steps that were actually taken to be read (such as when a step is taken with little to no arm movement), and some false positives will be registered based on arm movements that approximate the motions made when taking a step. However, contemporary wearables are still being seen as a way of getting individual information that can be a reasonably effective proxy of what had actually taken place and been experienced (Diaz et al., 2015). The same perspective limitations would apply to a wearable camera. While it is an approximation of what a person sees and hears, it is not necessarily aligned the same way as someone's eye gaze or ears. Furthermore, even if something is in the line of sight of a participant, it does not mean it was actually registered at the time. Thus, records from a wearable camera might capture more or less than what was actually 'seen' or 'heard' by a participant. Thus, caution must be exercised when making inferences from wearable camera footage.

1.3.1. Mobile learning theory and research wearables

The connection between wearable technologies and mobile learning theory has minimally been made (Lee, 2017), but is one that we wish to consider as part of this special issue. Mobile learning theory as an area for theory development has largely sought to elaborate on how potentials of mobile technology – canonically, handheld devices – are realized and how affordances of learning that takes place on personal and ubiquitous devices mediate new relationships between learners, educators, and content. Some of the main features of mobile learning that have been identified include the portability and mobility of learning, the ability to access content at any time that connectivity is accessible, the potential responsiveness of learning to cues from the local context and setting, means for communication to take place across devices and between individuals, and the ability for learning to be personalized (Crompton, 2013; Klopfer, 2017; Traxler, 2009). Wearables have not historically been included in mobile learning theory, likely because of their relative newness compared to more traditional portable computing devices. However, in her discussion of the need for theory related to mobile learning, Crompton (2013) has stated "one should consider [mobile learning] as the utilization of electronic devices that are easily transported and used anytime and anywhere" (p. 48). This makes wearables eligible for at least some speculative considerations as they relate mobile learning theory. They function like other mobile technologies in that they are easily transported and can be used anytime and anywhere. They can collect information and make information available for later learning and reflection. For instance, an activity tracker can store the number of steps taken in a day but also provide information on demand about how many steps were taken thus far or remind the wearer to be more active if they have been inactive for a period of time.

Yet while wearables can be eligible for consideration in mobile learning theory, there are points where the wearables used in this study, namely those that do not provide immediate user feedback or information, are an exception to current mobile learning theory. For instance, Sharples, Taylor, and Vavoula (2007) have sought to articulate a theory of mobile learning through an adaptation of the mediated relations commonly associated with the mediational triangle of cultural-historical activity theory (subject-mediating artifact-object). Typical mobile devices alter the relationship between a learner and a set of ideas or practices to be learned through their ongoing presence and ability to access and render information on demand. However, wearables that are largely meant to be operating in the background and outside of immediate awareness and that do not provide immediate feedback are limited in how they mediate immediate interactions and relations with the learning setting. That includes the wearables used in this study, which are meant to generate information that are examined post hoc. Thus, this more prominent example of mobile learning theory is largely silent on the role that the wearables in this study play. Similar limitations extend to other mobile learning theories such as those based on conversation theory (Sharples, Arnedillo-Sánchez, Milrad, & Vavoula, 2009) or transactional distance theory (Park, 2011). Some of this disconnect may reflect the positioning of some mobile technologies, in this case, wearables, as research instruments rather than learning instruments. Thus, it may be best to amend Crompton's above-mentioned criteria for mobile technology and consider mobile learning theory as applicable to electronic devices that are easily transported and used anytime anywhere while providing the ability to immediately capture and/or immediately return information to the user of the mobile technology. Doing so would make clear that immediate availability of information that is possible for some wearables is within the scope of mobile learning theory, whereas wearable research technologies that capture information that can only be examined much later speak to different considerations and fall outside of scope.

1.4. Maker learning theory

Finally, as this article examines engagement in Maker learning environment, we now turn to a discussion of the learning theories underlying Maker pedagogy. Proponents of Maker pedagogy often make an appeal to Papert's learning theory of constructionism, which stated that learning takes place especially felicitously in the process of constructing and sharing a public artifact of personal interest (Papert, 1980). Maker activities explicitly involve the creation of custom artifacts. In the Maker literature, the process of creating an artifact should "foster deep and agentive learning, in part because of the way in which it can generate reciprocal relationships among students and between students and teachers" (Bevan, 2017, p. 81). That is, traditional instruction with a teacher or facilitator leading a lesson or discussion is not supposed to be central to Maker pedagogy. Rather, peers, or even students and 'teachers', teach one another at different times depending on the task. Making is supposed to also place youth in contact with newer "expressive technologies" (Blikstein, 2013) that will allow for personal expression of interests and values, which should also promote youth engagement, presumably by building on existing interests. We would expect that customization activities that allow for personal embellishments, decorations, and modifications would promote engagement.

Also, participating in the Making community is supposed to support the development of specific dispositions (Ryan, Clapp, Ross, & Tishman, 2016) that are consistent with the "Maker Mindset" (Dougherty, 2013). The core values of the Maker mindset include being 1. playful 2. asset/growth oriented 3. failure positive and 4. collaborative (Martin, 2015). Those values are to be realized from the physical manipulation of materials as they are iteratively combined with one another to produce the culminating artifacts. Encountering mistakes and struggling with putting things together such that they 'work' eventually is supposed to engender a positive stance towards failure. Talking with peers and getting their input and assistance is supposed to support collaboration.

What is learned should include the values of the Maker mindset as well as science, mathematics, and engineering content and practices (Bevan, 2017). This is the result of disciplinary engagement that comes as a result of Making. Disciplinary engagements may come from acts of measurement and calculation in selecting and preparing materials, testing of material conductivity in a circuit while working with electronics, and iterative designing and building of artifacts as would be characteristic of engineering. Making an artifact may involve or lead to small scale experimentation and model construction. For instance, youth may make data collection instrumentation (e.g., a weather sensor to mount on a weather balloon) or may have to run iterative tests on sensors that are part of their culminating artifact to ensure that they are running correctly (e.g., a light sensor that provides input to determine which direction a robot moves). Some research has documented learning gains in these areas as a result of participating in Maker activities (Chu et al., 2017; Peppler & Glosson, 2013). More remains to be done to articulate the content learning and development of new practices from Making.

Finally, returning to psychological (rather than disciplinary) forms of engagement, we took the position in the introduction that the exact mechanisms for how engagement develops from Making have only been weakly specified. We noted that agentive positioning was supposed to be a key component, although how that was realized from types of Maker activities was undetermined. We might extrapolate also that repeated behavioral engagement in the form of revisiting Maker tasks is to be expected from the complexity of the materials being used (Keune, Peppler, & Wohlwend, in press). Cognitively, these tasks may be at a level of challenge such that they lead to a state of flow with individuals engrossed in the building of objects (Zollars, 2018). Synthesizing from the above points related to Maker learning theory, we can infer that characteristic Maker activities – collaborating, building objects with new materials, personal expression, doing activities that do not have a clear teacher leading a lesson - are supposed to lead to heightened engagement from youth. While those make intuitive sense, they still are subject to scrutiny as we do not know empirically if, for instance, a teacher led instructional presentation is indeed less likely to produce increased engagement for youth than an activity involving personal expression, such as customizing and decorating one's created artifact.

1.5. Research questions

In considering what is known in Maker education and what remains uncertain, we ask the following questions:

1. What activities in a Makerspace appear to elicit the largest number of instances of momentary engagement, operationalized as changes in electrodermal activity? As a specific instantiation of this question, we could ask: does an activity like teacher-led instruction lead to more or less momentary engagement than an activity like personal expression through artifact customization?
2. Given that Maker activities are, on the whole, assumed to elicit engagement, how consistent are the levels of engagement over time and across individuals? As a specific instantiation, should we expect fairly steady amounts of response across days in a Makerspace? And should we expect youths by and large to be uniformly engaged by specific activities?
3. How do analyses of momentary engagement, operationalized as changes in electrodermal activity, compare to a daily survey instrument to measure engagement? To what extent can changes in electrodermal activity be used as a proxy for engagement as self-reported?

2. Methods

2.1. Participants and setting

Participants came from an existing afterschool program hosted at a community Makerspace in the intermountain west region of the United States. The Makerspace was equipped with a range of equipment including 3D printers, programmable sewing machines, laser cutters, CNC machines, microcontrollers, robotics kits, laptop computers, and its own computer lab. Afterschool programs in the Makerspace were scheduled as multi-week project "camps" in which youth convened weekly for six-weeks. The camp organizers, as part of their own funding obligations and the personal interests from the Makerspace leadership team, had arranged for and designed camps exclusively for teen girls in the region at the time of our research study. The Makerspace was committed to making STEM accessible to youth through hands-on Making experiences, with most programs taking place during the summer and select camps being run during the academic school year. Data collection took place during the 2016–2017 academic year.

The camps that we observed were created by the Makerspace to include only young women. The target age group served by this makerspace was adolescents, which is common for community Makerspaces targeted toward youth. We recruited all young women participants who had already signed up to participate in two of the camps. All young women agreed to participate and provided parental consent and their own assent. The population was rural or from a small city near a research university. In the first camp, which emphasized Making around the topic of rocketry, there were 12 youth participants (Mean age = 12.27 SD = 1.27). The racial composition of that group was 10 White, 1 Latinx, and 1 Asian-American. In the second camp, which took place after the first one had completed its full six-week session, there were 13 youth participants (Mean age = 11.67 SD = 1.15). The racial composition of the second group was 11 white and 2 Latinx. The topic of the second camp was oriented toward the design and creation of custom laser-cut lanterns. Because these camps took place at the same Makerspace, some youth participants (N = 7, 6 White, 1 Latinx) participated in both camps. The demographics were consistent for the geographic area but lack some diversity that would be found in Makerspaces located in more urban locales. Adult facilitators, who were staff and volunteers at the Makerspace, also provided consent to be documented in this research study through video recording.

2.2. Materials

2.2.1. Engagement survey

We administered a survey instrument (Bathgate & Schunn, 2017) developed by another research group (Science Learning Activation Lab, 2015) to measure youth self-reported affective, behavioral, and cognitive engagement for each session. This instrument asked youth to rate their agreement to 8 statements on a 4 point Likert scale with response options being "YES!" "yes" "no" and "NO!". Sample statements included "During this activity, I felt excited" "…I felt bored" "…Time went by quickly" "I was focused on the things we were learning most of the time" and "…I was distracted". Bathgate and Schunn (2017) report on the reliability of this instrument (ranging from $\alpha = 0.71$ to 0.80, depending on which subscale) with similarly aged youth. The instrument had three items measuring the affective dimension of engagement, three items measuring the cognitive dimension, and two items measuring the behavioral dimension.

This survey was administered 4 days for the first camp (Rocketry emphasis) and for 5 days for the second camp (Laser-cut lantern emphasis). These were administered electronically on tablet devices provided by the Makerspace. For the first camp, one day was not surveyed because of internet connectivity problems (Week 3) and another involved launching rockets at a field site (Week 5) and inadequate time and infrastructure to distribute surveys. For the second camp, one day was not surveyed because the youth went on a tour of a research and design facility involved in aeronautics that is funded partially through government contracts. Security protocols for the facility forbade any recording equipment, so no data were collected that day.

2.2.2. Wearable devices

Each youth participant was provided with an Empatica E4 wristband (Fig. 1) to wear on their non-dominant wrist. The E4 is equipped to measure electrodermal activity through two electrodes that make contact with the forearm just below the wrist, recording EDA signals at 4 Hz (four times per second). As each session was approximately 90 min, that produced approximately 21,600 daily data points for each participant. The E4 has no active display nor interface except for a single button that must be pressed and held in order to ensure that a data recording session had begun. To end data recording, the button must be pressed down for several seconds. Data are transferred through a data transfer cradle attached to a separate computer.

In addition, each youth was provided with a GoPro Session camera (Fig. 1) that was worn on a chest mount configuration. The chest mount was determined to be the most physically comfortable worn arrangement for the camera (as opposed to a helmet or separate hat, see Stevens et al., 2016). The cameras were configured to obtain continuous firsthand participant video footage, with the inherent limitations described in Section 1.3 acknowledged.

2.2.3. Third-person video camera

For all but two sessions, a standing video camera operated by a member of the research team was used to obtain third-person video

Fig. 1. The Empatica E4 device (left) that continuously tracks EDA and the wearable GoPro Session camera that records video.

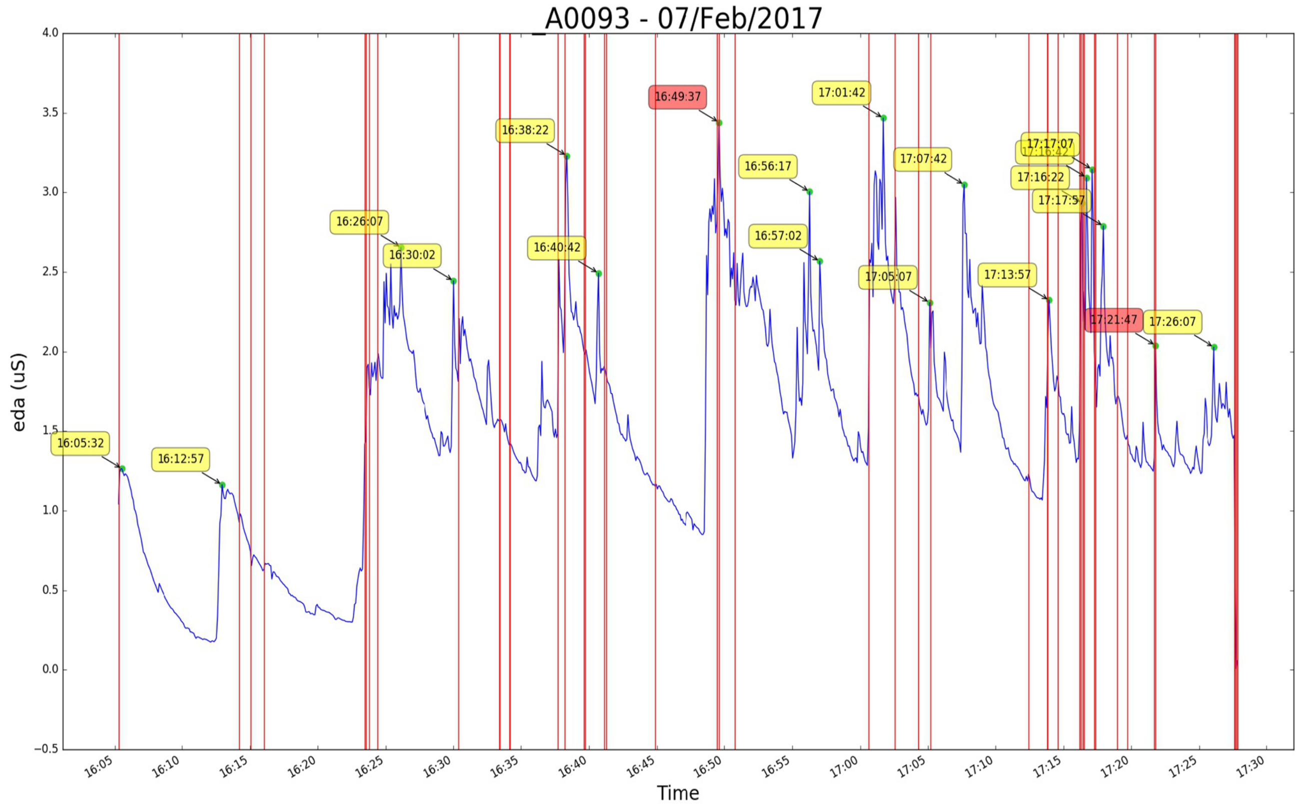

Fig. 2. EDA data obtained from one youth on one day of the second Maker camp. Peaks identified by our algorithm are highlighted. Peaks that were eliminated as potential data artifacts and their associated times are marked with the vertical line and darker shading in the timestamp boxes.

footage. One session during the first camp was missed due to a data transfer error. However, video footage from that day was represented from the set of all worn GoPro cameras from the participants. The other session, from the second camp, was not videorecorded as it was the day of the tour of the secured aeronautics facility.

2.3. Data collection procedures

Informed consent was obtained prior to the beginning of each camp. Outside of scheduled time, a phone interview was arranged with parental permission for a subset of randomly selected youth to administer some follow-up interviews. These were audiorecorded and transcribed.

A research team member was stationed at the Makerspace with assigned E4 devices and GoPro cameras. When each youth arrived, the research team member verified they were comfortable participating in the research study that day and helped them to put on the wearables and get them turned on. That team member also operated the standing video camera which primarily obtained a full group recording when possible. When youth changed rooms or moved to other locations, the research team member selected a subgroup to follow with the main camera.

At the end of each session, youth were provided with a tablet device that had the survey loaded for them to complete.

2.4. Data analysis

2.4.1. Survey analysis

The engagement survey data was scored following procedures described in Bathgate and Schunn (2017), with some items reversed and total scores computed for each individual. The maximum possible score was 32 (high scores reflecting higher self-rating for engagement). Average scores were also computed for each of the affective, behavioral, and cognitive subscales.

2.4.2. EDA peak detection

In traditional lab-based EDA research, data collection occurred in controlled laboratory conditions. Due to our person-in-context (Sinatra et al., 2015) approach to capturing engagement and due to researcher control of environmental conditions and stimuli being untenable, we have developed a different approach to EDA data analysis. As has been the tradition in lab-based studies, we focused on detecting peaks in EDA that suggested a sudden response to stimuli. Our assumption is that a peak corresponded to a change in stimulus in the preceding seconds that led to heightened arousal. We started with peak detection code made publicly available by the developers of the E4 (Taylor et al., 2015). We then modified the code to identify a peak as being when the relative change from one smoothed time interval (five seconds) to the subsequent time interval (the second five seconds) was greater than one positive standard deviation of each participant's daily mean. This was based on earlier work by Cain and Lee (2016) which suggested this technique could identify high arousal moments. The data were skewed positively (3.813). Because of the skew, this algorithm was assumed to identify less than the top 16% of peaks (which would be the value for an unskewed normal distribution). This reduced about 2.5 million data points to 1575 peaks (about 2.5% of the possible peaks from the entire corpus) (946 for the rocket camp, 629 for the lantern camp) that were greater than a single standard deviation for each participant each day, making it more conservative on what was considered noteworthy.

Once peaks were identified, potentially fallacious peaks were eliminated through use of the EDA explorer artifact detection algorithm developed by Taylor et al. (2015). In that algorithm, the developers used machine learning techniques to train the algorithm to identify potential noisy artifacts based on evaluations of sample data that were reviewed by two EDA experts. Co-existing peaks that occurred during time periods that the algorithm identified as potential noise or data artifacts were removed from further analyses, further reducing the set of peaks that were to be considered in our analysis (Fig. 2). That

Table 1
List of activity features identified from coding and iterative review of video data.

Feature	Description
Building on Existing Interest	The activity is explicitly oriented to an existing topical interest that individual youth have outside of the camp. For example, this would include times when a youth who likes soccer is asked to create images of soccer or talk about soccer
Digital Making	The activity emphasizes using digital tools to edit or modify some form of digital asset as would be common with media manipulation software. For example, this would include rescaling images and modifying them in graphics software
Dynamic Materials	Some material or equipment is used that produces some pronounced sensory stimuli, such as a flashing light, a loud noise, or a dramatic change in color. For example, this would include a model rocket launched into the sky
Interactive Instruction	The activity involves a facilitator, mentor, or tutor who is facilitating discourse around a topic of interest. The discourse should involve multiple speaker participants and may involve movement or demonstrations. For example, this would include a discussion of Newton's third law where a rolling chair is pushed into a nonmoving chair and youth predict what will happen
Personal Expression	The activity encourages some form of personal expression or aesthetic work, such as decorating an object or making something look an intended way. For example, this would include painting rockets in preferred colors and patterns
Peer Socializing	This activity provides time for youth to talk with one another about any topics that they wish that may not be relevant to the immediate task at hand, such as books they enjoy or things they noticed at school
Physical Making	The activity involves some form of assembly, creative construction, or tinkering (Bevan, 2017) with tangible objects
New Content	Some new disciplinary content, relative to what has been collectively discussed, is introduced and examined. For example, this would include a large group discussion about chemical reactions and property changes associated with different materials as they oxidize

reduced the number of peaks to 1029 (622 for the rocket camp, 407 for the lantern camp). An additional filter was implemented to remove days and peaks when participants for reasons unknown produced EDA signals that never exceeded one microSiemen (mS) either in base values or in peaks. That is, during our analysis, we identified one youth that almost consistently produced signals less than one mS, even when different devices were provided, suggesting she may simply be an individual with low overall skin conductivity. Two other participants from each camp were also identified that also consistently had signal values that did not exceed one mS. In addition to the participants that consistently displayed low peaks, three other participants had select days where their peaks did not exceed a one mS increase. Days on which participants did not produce peaks meeting our threshold were removed. This resulted in removal of data for 11 out of the 67 participant days from the rocket group and nine of the 77 participant days for the lantern group. It should be noted, that due to the conservative nature of our classification of peaks and the filtering from noise and low EDA signal values resulted in days in which some participants did not have any peaks for analysis by our measures. This reduced the number of peaks to 546 (324 for the rocket camp, 222 in the lantern camp). The remaining peaks were those that were included for coding and analysis, and are the peaks that were used for subsequent analysis and are referenced hereafter.

2.4.3. Video segment coding

2.4.3.1. Peak context coding. Time-stamps for all remaining peaks were identified and corresponding segments of footage obtained from each individual's GoPro camera were examined. In order to provide context to the activities associated with the peak, a 30-s window surrounding the peak (20 s prior to the peak and 10 s after the peak) was examined. When needed, other video were also used (e.g., the youth's camera was blocked by hair or an arm, or 3rd person or another participant's camera provided a better view of the situation).

Coding during this 30-s context window included codes developed from the data regarding who the individual was interacting with (*adult mentor, peers, computer, a combination, or none*), group level interaction (*large group, small group, one other individual, or self*) and activity type including (*Observing Making, Actively Making, Peer Play, Peer Socialization, Managing Materials, and Facilitated Instructional Interaction)*. These codes were adapted from activity types identified from initial review of data reported in Fischback and Lee (2017). For data from the first camp, two coders independently reviewed 56 video moments that were associated with EDA peaks and yielded $\kappa = 0.978$. For the second camp, two coders independently reviewed 45 video moments that were associated with EDA peaks and yielded $\kappa = 0.915$. The two coders then proceeded to independently split coding responsibilities for the remainder of all EDA peak data.

Each individual was also coded for their affective state based on review of facial expressions on the video at the time that a peak was detected. The scale used was a five point scale ranging from negative affect (a score of 1) to somewhat negative affect, to neither positive nor negative affect, to somewhat positive affect, to positive affect (a score of 5). Not all youth faces were visible even when running searches for specific individuals from video files recorded from the standing camera or other youths' wearable cameras. In the cases where no face was visible, no score was assigned. Two analysts independently coded 30 facial expressions and had agreement of $\kappa = 0.773$. Disagreements were resolved by discussion, and a single analyst then proceeded to code all affective states for youth based on facial expression at peak times identified during high EDA density peak clusters described in the next section.

2.4.3.2. Ridgeline analysis. In addition to coding individual peaks, we also identified time periods where there appeared to be a greater density of peaks across the attending youth and at least 25% of the youth exhibited peaks during that time period. This was done by preparing ridgeline plots with all youth EDA peak density aligned by time. The ridgeline plotting algorithm required at least three peaks from the same person during that day to produce a value in the density plot. Otherwise, that portion of the plot remained flat even though there were peaks in the data. Then a visual inspection for high density regions – where there was a youth's high density cluster of EDA peaks, represented by a 'hump' in the density data – aligned with other youth. The associated time was checked against a list of all peaks from all youth. A window of up to six minutes (roughly three minutes before and after the identified time in the ridgeline plot) was established. This time window was determined based on review of video data by two analysts who then jointly prepared a summary of the episode. From those, key features of the recorded activity were identified and assigned to that episode. The features were identified from the coded activities (Section 2.4.3.1) and from descriptive review of the video data. The purpose of this analysis was to identify what activities in a given week seemed to have the highest intensity of response (multiple EDA peaks) across multiple youth. Resultant activity features are summarized in Table 1.

3. Results

3.1. Engagement survey

Engagement survey descriptive statistics are provided in Table 2. For the days when data were obtained, the tendency was for all respondents to report each day was more engaging than not, with all individual scores across all days being greater than 16, and the lowest

Table 2
Engagement survey descriptives. Camp 1 is labeled as "Rocket". Camp 2 is labeled as "Lantern".

Week No.	Respondents	Mean Engagement Score (SD)	Mean Affective (SD)	Mean Behavioral (SD)	Mean Cognitive (SD)
Rocket 1	12	29.92 (3.90)	3.50 (0.63)	3.21 (0.66)	3.33 (0.59)
Rocket 2	9	27.22 (2.64)	3.78 (0.37)	3.06 (0.53)	3.26 (0.57)
Rocket 3	0	n/a	n/a	n/a	n/a
Rocket 4	10	25.60 (3.13)	3.50 (0.45)	2.85 (0.53)	3.13 (0.50)
Rocket 5	0	n/a	n/a	n/a	n/a
Rocket 6	11	27.82 (3.19)	3.73 (0.36)	3.18 (0.51)	3.42 (0.45)
Lantern 1	6	22.83 (3.92)	2.90 (0.50)	2.79 (0.76)	2.81 (0.60)
Lantern 2	0	n/a	n/a	n/a	n/a
Lantern 3	10	26.00 (2.00)	3.47 (0.45)	3.10 (0.66)	3.13 (0.28)
Lantern 4	9	26.00 (4.24)	3.48 (0.65)	3.06 (0.53)	3.15 (0.69)
Lantern 5	11	24.36 (3.88)	3.42(0.54)	2.59 (0.58)	2.97 (0.71)
Lantern 6	10	27.50 (2.59)	3.87 (0.23)	3.05 (0.50)	3.27 (0.52)

mean score on any recorded day being 22.83 (out of 32) and the lowest median being 24. Both of these were on the first day of the second camp. Subscore ratings were all rated on average greater than 2.5, suggesting the participants rated themselves as more engaged than not each day.

There was no significant difference across engagement survey responses across days during the first camp [$F(3, 38) = 0.83$, $p = 0.48$]. This held true for the subscales as well [Affective $F(3, 38) = 1.00$, $p = 0.401$, Behavioral $F(3, 38) = 0.879$, $p = 0.460$, and Cognitive $F(3, 38) = 0.555$, $p = 0.648$]. There was no significant difference across engagement survey responses across days during the second camp at the 0.05 level, although some may characterize it as trending [$F(4, 42) = 2.54$, $p = 0.054$]. We refer the reader to Wasserstein and Lazar (2016) for guidance on how to interpret the results and p values, as the alpha level of 0.05 can be inaccurately interpreted. When examining by subscales, it appeared that Affective showed a difference across days [Affective $F(4, 41) = 3.725$, $p = 0.011$, Behavioral $F(4, 41) = 1.376$, $p = 0.259$, Cognitive $F(4, 41) = 0.743$, $p = 0.568$]. The difference appeared to be from the first day of that camp that produced lower engagement scores, particularly affective scores. That day was predominantly focused on group introductions and a tutorial on how to use advanced graphics editing software.

3.1.1. Engagement survey responses and peaks per youth

We ran correlational analyses of the number of peaks per youth (computed from the values in Table 3) and the engagement total and subscales. This analysis was limited due to the small number of data points. It only included the number of girls whose EDA data was retained after our cleaning procedures (see Section 2.4.2). The correlation between the total engagement score from the survey instrument and peaks per person was $r = 0.671$ ($p = 0.048$). We found correlation values of $r = 0.330$ ($p = 0.386$) for affective engagement and peaks per person, $r = 0.625$ ($p = 0.072$) for behavioral engagement and peaks per person, and a value of $r = 0.645$ ($p = 0.061$) for cognitive engagement and peaks per person. Again, we refer the reader to Wasserstein and Lazar (2016) for guidance on interpreting these correlations. Our interpretation is that more peaks are correlated with higher engagement. Closer examination suggests more peaks are suggestive of higher cognitive engagement and behavioral engagement, but not affective engagement. We conjecture that is because electrodermal activity is most suited for responding to cognitive and behavioral dimensions because of its ties to psychophysiological arousal, but does not adequately differentiate between positive and negative affect, of which positive is most valued and measured by instruments such as the survey we used.

3.2. Distribution of peaks across camp days

Every participant except for those who were excluded for consistently low overall EDA signal had generated peaks at some point in their time at the Makerspace. In total, there were 324 peaks during the first camp and 222 peaks from the second camp (Table 3) that were included in our analysis.

Ridgeline plots were rendered of the data using the *ggplot* and *ggridge* packages in *R* to show the density of each youth's EDA peaks across each camp (Figs. 3 and 4). These plots show what proportion of that youth's EDA peaks throughout their camp appeared at different times. For example, if most of the participants EDA peaks took place around a certain time during week 2 and a different time during week 4, then there would be two density ridgelines in that person's row of the plot. For this particular visualization package, three data points from the same individual within a single day are required in order for a ridgeline to be shown. The plots are faceted by week (columns) and by participant (rows). The height of the ridgeline represents the proportion of their peaks that were at a certain time point. If an individual had 10 peaks with most of them appearing in the first five minutes and a few later on that same day, then a larger hump in the plot would appear in the first five minutes and a smaller hump or plateau would appear later on the plot. If an individual only had three peaks in close proximity to one another, there would be a single large hump at the time when those peaks had taken place.

The benefit of using ridgeline plot is to assist in identifying days and times when larger proportions of EDA peaks took place for individual youth. The ridgeline plot also shows profiles of individual participants over multiple days to allow for visual appraisal of similarity. If youth had similar responses, the overall contours of individual youth's ridgeline plots should be very similar to one another. They varied across youth. Ridgeline plots also allow for times when there were high densities of EDA peaks across multiple youths to be identified for subsequent review and analysis. The corresponding peaks for those participants was manually checked to see how many were the actual absolute number of peaks that took place at a given time. These

Table 3
EDA peaks from individuals with readable EDA data. Camp 1 is "Rockets" and camp 2 is "Lanterns".

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Total
# of Participants (Rockets)	5	8	12	9	10	9	
# of Peaks (Rockets)	44	63	48	53	44	72	324
# of Participants (Lanterns)	11	n/a	11	11	10	11	
# of Peaks. (Lanterns)	36	n/a	77	40	37	32	222

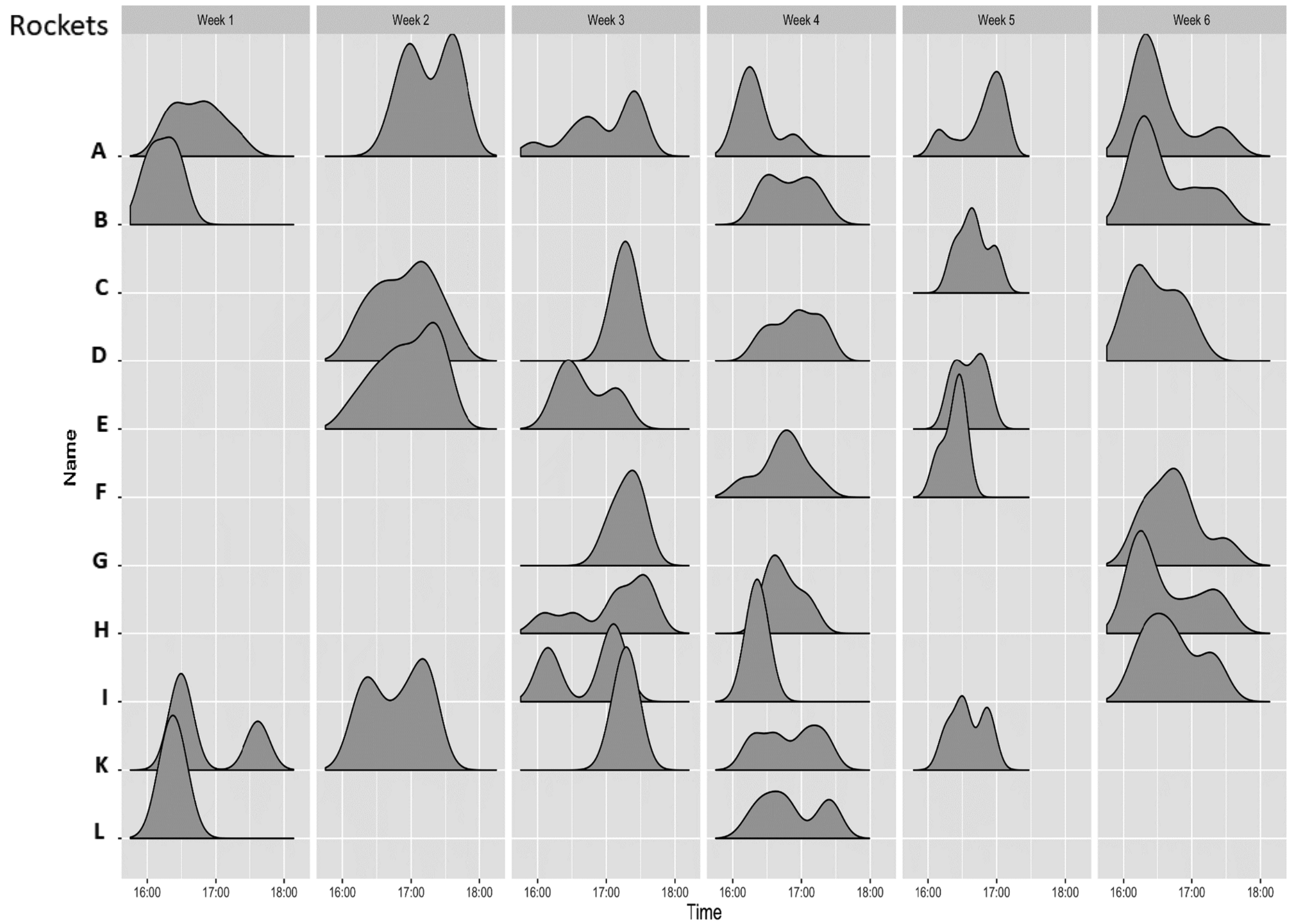

Fig. 3. Ridgeline plots of youth EDA peak densities during the rocketry Maker camp (Camp 1).

numbers of peaks are reported in Tables 4 and 5.

For camp 1, there were proportionally fewer density concentrations of EDA peaks during Week 1 (peer introductions and beginning to assemble rocket fins) and more during Weeks 2 (assembling rocket bodies) and 6 (comparing different potential rocket fuels), based on areas under the ridgelines. However, Weeks 3 (decorating rockets) and 4 (assembling rocket engine mounts and parachutes) had more participants with raised ridgelines, suggesting those weeks had elicited more frequent high density concentrations of EDA responses from more youth than other weeks. More youth responded on days when they were decorating rockets and assembling rocket engine mounts and parachutes. There was greater EDA response from select youth on days when they were assembling rocket bodies and comparing rocket fuels.

For camp 2, there were proportionally less EDA peak concentrations during Week 6 (installing light bulbs) and more during weeks 4 (laser cutting lamp components) and 5 (assembling lamp components). Week 3 (selecting and editing custom images for lantern decoration) had the most participants with raised ridgelines, suggesting that week had more frequent EDA responses from more youth than other weeks. From this, it would appear that weeks with customization and working with expressive technologies tend to yield EDA response across more youth than other activities.

3.3. Features of high peak density moments

Looking across ridgeline plots helps to determine what weeks yielded EDA responses across multiple youth. That is, we could determine which weeks had EDA responses for the majority of the participants. It can also help identify specific times during each week that exhibited a larger concentration of peaks. That is, we could determine what activities during those weeks contributed to a sizable proportion of EDA responses. This is a different analysis in that the entire week may have elicited a response from youth, but it could have been that the activities during that week were so varied that only one youth was showing signs of engagement in a given window of time according to EDA response. For instance, personalization and customization could have peaks scattered throughout the week as different youth made customizations to their constructed artifacts at different moments. Looking at high density moments that co-occur – precise times when multiple youth were simultaneously exhibiting multiple EDA peaks – would tell us more about what specific activity during the week was potentially engaging to multiple youth simultaneously. We refer to those times as high EDA peak density moments.

For the first camp, 13 high EDA peak density moments were identified. For the second camp, 9 high EDA peak density moments were identified. These are summarized in Tables 4 and 5 with their date and time of occurrence, as well as the number of peaks within those times and the number of youth who produced EDA peaks. In order to gain a further insight concerning participant affect during peaks, multimodal analysis of peaks also included the results of affective coding from facial expressions in Tables 4 and 5. On high EDA peak density moments that had NA for standard deviation or for the mean, there was only one or no faces visible at any camera angle respectively to make a judgment. Most high EDA peak density moments had mean affective scores greater than 3.

The most frequently appearing feature across high EDA peak density moments was peer socializing (13 instances). The next most common were interactive instruction and physical Making (7 instances each). Dynamic materials and digital Making each had 4 instances. The least common features were new content (3 instances), personal expression

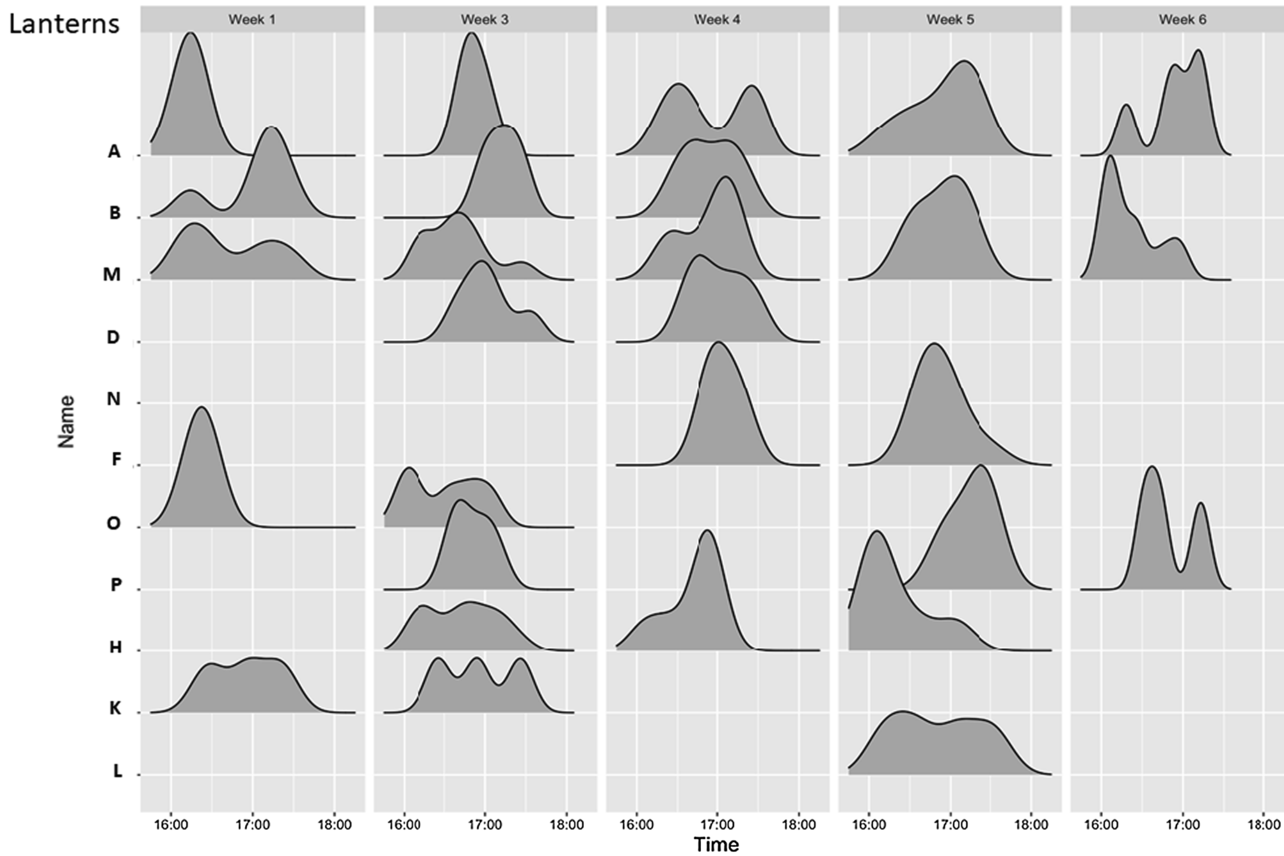

Fig. 4. Ridgeline plots of youth EDA peak densities from the laser-cut lantern Maker camp (Camp 2).

(1 instance), and building on existing interest (1 instance).

This suggests that the activities that were structured to allow for more peer socializing tended to show the highest levels of engagement. Those were almost always during digital Making or physical Making. While that may seem contradictory with respect to youth being engaged as they were socializing when they were given time to work on Making, our contention is that during those activities, youth were actually simultaneously engaged with their social conversations and interactions. That is what they were engaging with while also doing some of their Making work. It would be akin to a knitting circle gathering and talking while knitting. The Making, we suspect, provided a context and space for additional conversations to take place.

What is also surprising is that there were a relatively high number of high EDA peak density moments when there was instruction taking place. The instruction was interactive in that it was not a strict lecture format, but it was clearly a time when an adult facilitator was leading a group and the discussion, much as what we would see a skilled teacher do in a classroom. This applied to whole group instruction (more common in the rocket camp) and to instruction in small groups (more common in the lantern camp) but involved an adult expert in the room providing the instruction. This runs counter to the claim that it is the flattened power structure for who does the work of teaching that is especially engaging about Making. Perhaps instruction can be more intensely engaging for multiple youth simultaneously, while less structured and reciprocal instructional interactions – as can be realized during peer socializing – are engaging for youth on a more ad hoc basis. That is, to get a response from more youths at the same time, adult-led instruction can be engaging for youth in a Makerspace. To get engagement throughout a longer period of time, less structured time with more opportunity for peer socializing appears to be engaging.

4. Discussion

4.1. Summary and theoretical implications

Our analyses yielded several observations based on the assumption that EDA response has the potential to inform us about engagement. First, there appears to be some correlation between EDA response and overall engagement as measured by survey. More specifically, the correlation appears to be between cognitive engagement and behavioral engagement. Affective engagement does not seem to be meaningfully correlated with EDA responses. We believe this is because electrophysiological response is more suited to capture heightened arousal and immediate attentional processes that would be associated with cognitive and behavioral dimensions of engagement. EDA data alone would not differentiate between positive and negative affect, as both valences should yield an arousal response. Positive affect is what was measured in the survey instrument, and thus had weaker correlations. More data would need to be collected in the future to strengthen those claims. However, considering that EDA has been effectively used as a measure of cognitive load in the past and that negative affective response can be detected with EDA, this claim seems plausible. Thus, we claim that EDA is a resource for detecting engagement, particularly the cognitive and behavioral dimensions, in situ.

This means that our exploratory approach for taking a person-in-context approach for detecting moments of engagement has demonstrated plausibility. Where we note some challenges with engagement theory is that there is the inherent assumption in some instruments and the general treatment of engagement that engagement must have positive affect associated with it. As we had discussed in the introduction, and as suggested from some of our correlations, we may have moments of engagement that do not have positive affect. Regardless, they may have cognitive and behavioral engagement represented. We suggest

Table 4
Rocket (Camp 1) high EDA peak density moments.

Date	Time	No. of EDA Peaks	No. of Youth with Peaks	Percent of day's EDA peaks	Mean Observed Affective Score (SD)	Summary	Key Features of Activity
10/11/16	16:24–16:32	11	4	25.00%	4.17 (0.72)	Facilitator is leading interactive discussion explaining physics of rocketry, demonstrating with youth	Interactive instruction New content
10/18/16	16:25–16:30	5	3	7.94%	2.50 (0.58)	Youth are assembling rockets, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
10/18/16	17:01–17:05	8	3	12.70%	3.38 (0.52)	Youth are assembling rockets, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
10/25/16	17:09–17:13	8	6	16.67%	3 (na)	Youth are decorating their rockets with paint and talking about other activities	Personal Expression Peer Socializing
10/25/16	17:19–17:25	7	4	14.58%	3.25 (0.50)	Youth are decorating their rockets with paint and talking about other activities	Personal Expression Peer Socializing
11/1/16	16:23–16:25	8	6	15.09%	3.75 (1.17)	Youth are assembling rockets, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
11/1/16	16:33–16:38	7	5	13.21%	3.29 (1.11)	Youth are assembling rockets, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
11/8/16	16:23–16:29	5	3	11.36%	5.00 (na)	Preparing and observing first pair of rocket launches	Dynamic materials
11/8/16	16:46–16:50	5	3	11.36%	5.00 (0.00)	Observing rocket launches and free time conversing with peers	Dynamic materials Peer Socializing
11/15/16	16:10–16:16	18	6	25.00%	3.94 (0.87)	Facilitator is leading interactive group discussion and explaining differences in rocket fuels	Interactive instruction New content
11/15/16	16:24–16:29	7	3	9.72%	3.71 (0.95)	Facilitator is leading interactive group discussion and explaining differences in rocket fuels	Interactive instruction New content
11/15/16	16:58–17:03	7	6	9.72%	3.00 (0.00)	Facilitator is lighting rocket fuels, youth are talking freely with each other	Dynamic materials Peer Socializing
11/15/16	17:20–17:26	8	6	11.11%	4.00 (na)	Facilitator is lighting rocket fuels, youth are talking freely with each other	Dynamic materials Peer Socializing

Table 5
Lantern (Camp 2) high EDA peak density moments.

Date	Time	No. of EDA Peaks	No. of Youth with Peaks	Percent of day's EDA peaks	Mean Observed Affective Score (SD)	Summary	Key Features of Activity
1/10/17	16:14–16:18	12	8	33.33%	4.00 (0.76)	Beginning of camp, youth are introducing themselves to one another	Peer Socializing
1/24/17	16:15–16:20	9	3	11.69%	na	Youth are searching for images to use for their laser-cut lantern faces	Building on existing interest
1/24/17	16:35–16:40	8	3	10.39%	3.00 (0.00)	Youth work on editing selected images, youth with peaks are receiving one-on-one help from a mentor or are adjacent to someone who is receiving help	Interactive instruction Digital Making
1/24/17	16:50–16:56	10	5	12.99%	na	Youth work on editing selected images, youth with peaks are receiving one-on-one help from a mentor or are adjacent to someone who is receiving help	Interactive instruction Digital Making
1/31/17	16:24–16:29	8	5	20.00%	2.80 (0.45)	Youth work on editing selected images, youth with peaks are receiving one-on-one help from a mentor or are adjacent to someone who is receiving help	Interactive instruction Digital Making
1/31/17	16:56–17:01	5	3	12.50%	3.33 (0.58)	Youth work on editing selected images, youth with peaks are receiving one-on-one help from a mentor or are adjacent to someone who is receiving help	Interactive instruction Digital Making
2/7/17	16:34–16:40	6	3	16.22%	5.00 (0.00)	Youth are assembling lanterns, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
2/7/17	17:01–17:06	7	5	18.92%	4.00 (1.00)	Youth are assembling lanterns, helping each other figure out pieces, and having conversations about other activities	Physical Making Peer Socializing
2/14/17	16:42–16:48	6	5	18.75%	4.00 (0.82)	Youth are assembling lanterns, testing them, playing with them, and having conversations about other activities	Physical Making Peer Socializing

that in future discussions of engagement, the presumption that engagement is associated with positive affect be made explicit. Some promising research could still be done on moments that are engaging but not associated with positive affect. Those may not lead to the development of interest in the long term, but may lead to response and attention from participants. We may want to document those in future research on engagement.

Second, according to the data analyzed here, the days that more youth were engaged were ones that involved decorating, personalization, and assembly. This seems consistent with the notion that Making is engaging because of opportunities for personal expression and working with expressive technologies, as learning theory related to making would suggest. However, what seems to be commonly engaging for youth appears to be the opportunity to socialize with one another while they are personalizing and assembling materials. It may be that the combination of both interacting with peers and working on an assembly or decorative task is what is engaging. We note that this is not exclusively off-task conversation. Sometimes, the conversation was about other activities or topics of interest, such as books that they have read. Sometimes the talk was collaborative and supportive and related to the task at hand, as was this sample exchange from the rockets camp when one girl was getting help from another while assembling their rocket.

Youth 1. Can you help me?
Youth 2. Sure.
Youth 1. It's not perfect.
Youth 2. Here how about this. I'll hold this straight.
Youth 1. I need glue. Hey can I borrow that glue? If you're done. Thank you!
Youth 2. Okay, I'll hold it for you, and you glue. Alright, go fast. So it doesn't. Alright, there. Put it like here.
Youth 1. Oh hey, thanks for helping me.

Those transactions were part of the broad range of conversations that took place during peer socializing and were where engagement across peer socializing activities appeared in conjunction with digital and physical Making activities. They were sporadic throughout a given day, but represent moments of engagement that appeared in the data. This suggests that there is an important social dimension to learning in Maker experiences that should be made more prominent in any theory explaining what makes Making engaging.

To get more youth to be engaged *at the same time*, expert-facilitated instruction seems to be effective. One of the challenges that this exploratory study presents to the literature is the way in which a formal instructor and centralized instruction can still be engaging in a Makerspace. While the reciprocity of teaching and learning roles may be characteristic of learning in a Makerspace (Sheridan et al., 2014), there does seem to be some benefit in terms of triggering engagement from limited expert-led instruction. This is a refinement to ideas about limited instruction in Maker learning activities. Centrally-led instruction can cue simultaneous multi-participant engagement, and that may be desirable depending on one's aims.

However, considering also how Making is beginning to appear in more formal learning settings (e.g., Chu et al., 2017), the results of this analysis suggest caution for overly structuring classroom-based maker experiences. For instance, one can imagine that a teacher might structure a Making-inspired lesson as a quiet activity to be done individually at student desks where talking with peers is not permitted or highly discouraged. They might take the assertion that centralized instruction can cue engagement from this work as justification. However, based on the results of the current analysis, overuse of that would appear to be ill-advised. Socializing and working with one another while Making seems to be part of what made Making engaging for the youth in this study. Centralized instruction has its time and place in Maker learning activities, but so do more open ended tasks where peer socialization is a key component. A balance should be struck in Maker learning activities.

As far as mobile learning theory is concerned, we have shown how a mobile technology – namely, wearables that obtain EDA data and video footage – can be useful for research purposes yet pose challenges for mobile learning theory. Our suggestion had been to better circumscribe what technologies are considered central for mobile learning theory, as not all mobile technologies provide the same affordances for the learner. In this case, the information that was available was available post hoc and to third parties (researchers) rather than the learners themselves. There are cases when wearables are useful learning technologies because the data they generate returns to them (Lee, 2013). However, that was not the case here, where wearables served as a platform for conducting academic research.

4.2. Limitations and challenges

As an exploratory study, we are aware of several limitations. First, while thousands of data points were generated, the number of participants was relatively small. This is due to the enrollment in the Maker programs that we had used as research sites. The research population was all female, which also was a limitation that came with the research sites. It may be that a mixed group of genders would yield different responses, or that males would find different aspects of Making engaging than what we documented here. Looking at and comparing with a more diverse sample remains work for the future. Furthermore, technical difficulties and security restrictions limited how much data we were able to obtain, especially from daily surveys. Thus, there are limits to generalizability from this study.

Another limitation is that we were presumably capturing moments of engagement whereas engagement could last for longer periods of time than our instrumentation could capture. Examining peaks from EDA data captured engagement as it was triggered, we believe, and lasted for a few seconds. This technique does not capture sustained engagement or sustained interest. It may be that youth were overall more engaged than we present here because they were engrossed in their tasks. They may not have exhibited a psychophysiological response though. Other methods would need to be recruited to make that determination.

One such method would be to get retrospective reports of self-identified engagement from youth at shorter time intervals. This could be done by administering the surveys more often than once per weekly meeting, or it could be done with a self-rating of engagement. In fact, we attempted the latter approach, but concluded that our participating youth were not well equipped to provide retrospective self-reports about what was more or less engaging in a given day. When we tried to get retrospective accounts of engagement from youth to validate what we were getting from wearable devices, we found that youth accounts of what happened and when it happened was either too vague (e.g., "I was into most of it. It was fine.") or incongruent with what we had recorded in video data. Longstanding psychological research on retrospective self-ratings suggests that we can be easily swayed by some recalled events and specific features of experiences that may not reflect their summative qualities (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Given that and the data quality, we did not pursue further analysis on the participants' qualitative retrospective accounts of what was or was not engaging at different times.

An alternative approach could have been to utilize experience sampling techniques in which youth were asked at random intervals to report what they are doing and their levels of engagement. Experience sampling shows good potential (Xie, Heddy, & Greene, 2019), although it may require more time with youth than the duration of time we had available to generate sufficient data. Furthermore, in close confines and with this age group, we were concerned that youth would be distracted by each other responding to prompts to provide samples at different times.

What is also a concern, based on the engagement survey responses, is that self-appraisals may be inflated without more deliberate effort to help youth recalibrate and classify their levels of engagement. Youth in this study were already willing to be part of a Maker camp experience outside of school time, which may bias them to be on the higher end of many existing instrument scales. Their engagement scores on the engagement survey were relatively high, with few youths responding low engagement. This could have been because the activities were more engaging than not, but it could also be due to the self-selected population that was studied. Still, more research comparing trade-offs with different sampling techniques is needed whether it is with comparable or different populations.

A challenge with the approach described here is that this methodological approach has also been time consuming to analyze. Thousands of data points were generated and hundreds of hours of video footage were produced. Analytically, this required careful storage systems and cross referencing video footage from several angles and perspectives. As computational techniques and more standards are established, this may become an easier task for future analysts. However, a substantial amount of time making sense of human activities and interactions was required to come up with the classifications that we did. Distributing and relying on more likert scale surveys would have its own challenges, but analyses of that kind of data would likely have more established procedures that could expedite the process compared to what we had done.

4.3. EDA wearables as a research tool

Circling back to the special issue theme of mobile technologies, we believe we have some practical lessons to shed on wearables as a class of mobile technology. First, we have used wearables as a research instrument rather than as a tool for pedagogy, which expands on Lee & Shapiro's theory of wearables as providing forms of learner support (2019). Lee & Shapiro argued that an emerging form of learner support was the ability to obtain bodily records but the assumption was that the learner would be subsequently examining those records. The usefulness of bodily records from wearables goes beyond providing feedback and supporting the learning of the wearer and instead serves as part of a platform for doing research. Specifically, EDA wearables have some initial promise. They may not capture all that researchers value as being engaging, but they can provide some pointers for future inquiry. If the correlations that we saw are indeed valid and hold up to future research, then we have uncovered a way of getting data about youth engagement in a learning activity without requiring surveys or interruptions. However, we do wish to reiterate that EDA wearables do not work effectively on all individuals. Some youth produced data that did not produce enough of a signal for us to analyze. Others should recognize that they offer some unique features, such as their ability to obtain large amounts of data, but they also presume particular indicators of some phenomenon (i.e., EDA peaks are markers of engagement) that may not be universally accepted. For instance, Di Lascio et al. (2018) used overall skin conductivity as a measure of engagement rather than peaks in EDA data. That is a different way of computing engagement from EDA signals than what we had done here.

For EDA data to be actionable beyond stating a magnitude of engagement, richer accounts of context are still valuable and necessary. We tried to obtain those richer accounts with wearable cameras. At the same time, those richer accounts were costly to analyze. There were also occasions when the cameras, as mounted, did not provide useful footage because something (e.g., an arm, a table edge) was blocking the view of the camera lens. Moreover, the video segmenting system of the wearable cameras, which automatically created clips segmented in 12 min intervals, created additional work for us to stitch data back together. We had begun this study assuming that aligning video to EDA data would be a straightforward task, but it turned out to be a time-intensive one because of how file formats were created and stored on different wearable devices.

Also, EDA wearables still seem limited in detecting affective states. Our correlational analyses suggest affect is not detected in a significant way. As engagement is a multidimensional construct that has an affective component, this is a concern for future work. We have tried to work around this by using a survey instrument that had some affective components and by examining facial expression data to infer affective state. However, these were compensatory actions for a shortcoming in relying heavily on EDA to measure engagement. Perhaps the most promising direction is to use multimodal technologies that can take EDA as one data stream with others that can detect affect (D'Mello et al., 2017).

5. Conclusions

Regarding engagement in making, we have thus begun to make some headway in understanding if and how Maker learning activities support engagement. To revisit the research questions, the activities that seemed to elicit the most momentary engagement from youth in a makerspace were those that involved peer socializing, physical making, and interactive instruction. The opportunity to participate in activities related to personal expression was also engaging, but there were few high EDA peak density moments, suggesting the engagement for any given individual during personal expression activities is more sporadic. For Maker pedagogy, we see more importance of expert-led instruction than has been highlighted in the literature and the importance of peer socializing as part of what makes Making activities engaging.

Across weeks, there were some days that had more EDA response across youth, such as occasions of personal expression through decorating rockets and assembling engine mounts and parachutes for the rocket camp and personal expression through selecting and editing images to place on their custom lanterns for the lantern camp. From this, we can infer that more youth will have a response on days when they can pursue personal expression but not at the same time as each other.

Finally, we had some correlational results that suggest the number of EDA peaks per person correlates moderately with self-reported survey scores of engagement. More specifically, we have some suggestions of moderate correlations between peaks per person and cognitive and behavioral engagement. Affective engagement does not seem to be meaningfully correlated with EDA peaks. Methodologically, it suggests that counting EDA peaks has some potential for detecting some features of learning activities and environments that are momentarily engaging. Future work will be necessary to bolster or refute these early findings.

Acknowledgments

Thanks go to Aditya Chandel, Diamond Dang, Kourtney Schut, and Kyle Lam for their assistance with data collection and data processing. Thanks also go to Kevin Reeve and Dallin Graham for helping provide access to the research site. Helpful comments that improved the manuscript were provided by the anonymous reviewers and by Matt Bernacki. This work was supported in part by funding from the National Science Foundation under Grant No. CNS-1623401. The opinions expressed herein are those of the authors and do not necessarily reflect those of the National Science Foundation.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2019.101789.

References

Ainley, M. (2012). Students' interest and engagement in classroom activities. *Handbook of research on student engagement* (pp. 283–302). Boston, MA: Springer.

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology, 87*(4), 559.

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069.

Azevedo, R., Taub, M., Mudrick, N., Farnsworth, J., & Martin, S. A. (2016). Interdisciplinary research methods used to investigate emotions with advanced learning technologies. In M. Zembylas, & P. A. Schutz (Eds.). *Methodological advances in research on emotion and education* (pp. 231–243). Cham: Springer International Publishing.

Barton, A. C., & Tan, E. (2018). A longitudinal study of equity-oriented STEM-rich making among youth from historically marginalized communities. *American Educational Research Journal, 55*(4), 761–800. https://doi.org/10.3102/0002831218758668.

Bathgate, M., & Schunn, C. (2017). Factors that deepen or attenuate decline of science utility value during the middle school years. *Contemporary Educational Psychology, 49*, 215–225. https://doi.org/10.1016/j.cedpsych.2017.02.005.

Bernacki, M. L., & Walkington, C. (2018). The role of situational interest in personalized learning. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000250.

Bevan, B. (2017). The promise and the promises of making in science education. *Studies in Science Education, 53*(1), 75–103. https://doi.org/10.1080/03057267.2016.1275380.

Blikstein, P. (2013). Digital fabrication and 'making' in education: The democratization of invention. In J. Walter-Herrmann, & C. Büching (Eds.). *FabLabs: Of machines, makers and inventors* (pp. 203–222). Bielefeld, Germany: Transcript Publishers.

Cain, R., & Lee, V. R. (2016). Measuring electrodermal activity to capture engagement in an afterschool maker program. In P. Blikstein, M. Berland, & D. A. Fields (Eds.). *Proceedings of FabLearn 2016: 6th annual conference on creativity and making in education* (pp. 78–81). Stanford, CA: ACM.

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*(3), 253–264. https://doi.org/10.1016/S0167-8760(01)00147-7.

Chu, S. L., Schlegel, R., Quek, F., Christy, A., & Chen, K. (2017). *'I Make, Therefore I Am': The effects of curriculum-aligned making. Proceedings of the 2017 CHI conference on human factors in computing systems, Denver, Colorado, USA*.

Cordova, J. R., Sinatra, G. M., Jones, S. H., Taasoobshirazi, G., & Lombardi, D. (2014). Confidence in prior knowledge, self-efficacy, interest and prior knowledge: Influences on conceptual change. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2014.03.006.

Crompton, H. (2013). Mobile learning: New approach, new theory. In Z. Berge, & L. Y. Muilenburg (Eds.). *Handbook of mobile learning* (pp. 47–57). Routledge.

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*(2), 107–126.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.). *Handbook of psychophysiology* (pp. 159–181). 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. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(3), 1–21. https://doi.org/10.1145/3264913.

Diaz, K. M., Krupka, D. J., Chang, M. J., Peacock, J., Ma, Y., Goldsmith, J., ... Davidson, K. W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology, 185*, 138–140. https://doi.org/10.1016/j.ijcard.2015.03.038.

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123. https://doi.org/10.1080/00461520.2017.1281747.

Dougherty, D. (2013). The maker mindset. In M. Honey, & D. Kanter (Eds.). *Design, make, play: Growing the next generation of STEM innovators* (pp. 7–11). 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*(4), 399–483.

Fischback, L., & Lee, V. R. (2017). How time gets used in afterschool maker programs. In P. Blikstein, M. Berland, & D. A. Fields (Eds.). *Proceedings of FabLearn 2017: 7th annual conference on creativity and making in education* (pp. 15–18). Stanford, CA: ACM. https://doi.org/10.1145/3141798.3141810.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059.

Grafsgaard, J. F., Wiggins, J. B., Vail, A. K., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2014). The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring. *Proceedings of the 16th international conference on multimodal interaction* (pp. 42–49). ACM. https://doi.org/10.1145/2663204.2663264.

Gruzelier, J. H., & Venables, P. H. (1972). Skin conductance orienting activity in a heterogeneous sample of schizophrenics. *Journal of Nervous and Mental Disease, 155*(4), 277–287. https://doi.org/10.1097/00005053-197210000-00007.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4.

Hsu, Y.-C., Baldwin, S., & Ching, Y.-H. (2017). Learning through making and maker education. *TechTrends, 61*(6), 589–594. https://doi.org/10.1007/s11528-017-

0172-6.
Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x.
Kang, S., Norooz, L., Oguamanam, V., Plane, A. C., Clegg, T. L., & Froehlich, J. E. (2016). *SharedPhys: Live physiological sensing, whole-body interaction, and large-screen visualizations to support. Proceedings of the 15th international conference on interaction design and children, Manchester, United Kingdom*.
Keune, A., Peppler, K. A., & Wohlwend, K. E. (2019). Recognition in makerspaces: Supporting opportunities for women to “make” a STEM career. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2019.05.013 (in press).
Klopfer, E. (2017). Mobile learning. In K. Peppler (Ed.). *SAGE encyclopedia of out-of-school learning* (pp. 500–502). Thousand Oaks, CA: SAGE.
Latulipe, C., Carroll, E. A., & Lottridge, D. (2011). *Love, hate, arousal and engagement: exploring audience responses to performing arts. Proceedings of the SIGCHI conference on human factors in computing systems, Vancouver, BC, Canada*.
Lee, V. R. (2013). The Quantified Self (QS) movement and some emerging opportunities for the educational technology field. *Educational Technology, 53*(6), 39–42.
Lee, V. R. (2017). Mobile devices. In K. Peppler (Ed.). *The SAGE encyclopedia of out-of-school learning* (pp. 496–498). Thousand Oaks, CA: SAGE Publications.
Lee, V. R., & Shapiro, R. B. (2019). A broad view of wearables as learning technologies: Current and emerging applications. In P. Diaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.). *Learning in a digital world – Perspectives on interactive technologies for formal and informal education* (pp. 113–133). Singapore: Springer.
Lee, V. R., Drake, J. R., & Thayne, J. L. (2016). Appropriating quantified self technologies to improve elementary statistical teaching and learning. *IEEE Transactions on Learning Technologies, 9*(4), 354–365. https://doi.org/10.1109/TLT.2016.2597142.
Linnenbrink-Garcia, L., Patall, E. A., & Messersmith, E. E. (2013). Antecedents and consequences of situational interest. *British Journal of Educational Psychology*. https://doi.org/10.1111/j.2044-8279.2012.02080.x.
Lyons, L. (2015). Exhibiting data: Using body-as-interface designs to engage visitors with data visualizations. In V. R. Lee (Ed.). *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 185–200). New York, NY: Routledge.
Martin, L. (2015). The promise of the maker movement for education. *Journal of Pre-College Engineering Education Research (J-PEER), 5*(1), 4.
Matsumoto, R. R., Walker, B. B., Walker, J. M., & Hughes, H. C. (1990). Fundamentals of neuroscience. In J. T. Cacioppo, & L. G. Tassinary (Eds.). *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 58–112). New York, NY, US: Cambridge University Press.
Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist, 50*(1), 31–42. https://doi.org/10.1080/00461520.2015.1004068.
Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2007.03.001.
Naveteur, J., & Baque, E. (1987). Individual differences in electrodermal activity as a function of subjects' anxiety. *Personality and Individual Differences, 8*(5), 615–626. https://doi.org/10.1016/0191-8869(87)90059-6.
Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). *Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. Paper presented at the proceedings of the 24th Australian computer-human interaction conference, Melbourne, Australia.*.
Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* New York, NY: Basic Books.
Park, Y. (2011). A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types. *The International Review of Research in Open and Distributed Learning, 12*(2), 78–102.
Peppler, K., & Glosson, D. (2013). Stitching circuits: Learning about circuitry through E-textile materials. *Journal of Science Education and Technology, 22*(5), 751–763. https://doi.org/10.1007/s10956-012-9428-2.
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*(5), 1243–1252. https://doi.org/10.1109/tbme.2009.2038487.
Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist, 50*(1), 58–69. https://doi.org/10.1080/00461520.2014.999920.
Renninger, K. A., & Hidi, S. E. (2016). *The power of interest for motivation and learning.* New York: Routledge.
Ryan, J. O., Clapp, E. P., Ross, J., & Tishman, S. (2016). Making, thinking, and understanding: A dispositional approach to maker-centered learning. In K. Peppler, E. Halverson, & Y. Kafai (Eds.). *Makeology* (pp. 29–44). New York, NY: Routledge.
Ryu, S., & Lombardi, D. (2015). Coding classroom interactions for collective and individual engagement. *Educational Psychologist, 50*(1), 70–83. https://doi.org/10.1080/00461520.2014.1001891.
Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2017). Enhancing student models in game-based learning with facial expression recognition. *Proceedings of the 25th conference on user modeling, adaptation and personalization* (pp. 192–201). ACM. https://doi.org/10.1145/3079628.3079686.
Science Learning Activation Lab (2015). Measures technical brief: Engagement in science learning activities version 3.1 August 2015. Retrieved from http://www.activationlab.org/wp-content/uploads/2015/09/Engagement-Report-20150831.pdf.
Setz, C., Arnrich, B., Schumm, J., Marca, R. L., Tr, G., #246, ... Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *Transactions Information Technology in Biomedicine, 14*(2), 410–417. https://doi.org/10.1109/titb.2009.2036164.
Sharples, M., Arnedillo-Sánchez, I., Milrad, M., & Vavoula, G. (2009). Mobile learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.). *Technology-enhanced learning: Principles and products* (pp. 233–249). Dordrecht: Springer Netherlands.
Sharples, M., Taylor, J., & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews, & C. Hathornthwaite (Eds.). *The SAGE handbook of E-learning research* (pp. 221–247). Thousand Oaks, CA: SAGE.
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*(4), 505–531.
Sherin, M. G., Russ, R. S., Sherin, B. L., & Colestock, A. (2008). Professional vision in action: An exploratory study. *Issues in Teacher Education, 17*(2), 27–46.
Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). *Galvanic skin response (GSR) as an index of cognitive load. Paper presented at the CHI '07 extended abstracts on human factors in computing systems, San Jose, CA, USA*.
Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924.
Stevens, R., Jona, K., Penney, L., Champion, D., Ramey, K. E., Hilppö, J., ... Penuel, W. (2016). *FUSE: An alternative infrastructure for empowering learners in schools. 12th International conference of the learning sciences: Transforming learning, empowering learners ICLS 2016*. International Society of the Learning Sciences (ISLS).
Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). *Automatic identification of artifacts in electrodermal activity data. International Conference of the IEEE engineering in medicine and biology society (EMBC)*. Milan, Italy: IEEE.
Traxler, J. (2009). Current state of mobile learning. In M. Ally (Vol. Ed.), *Mobile learning: Transforming the delivery of education and training: Vol. 9-24*. Edmonton, AB: AU Press.
Umphress, J., & Sherin, B. (2015). The body as viewfinder: Using wearable cameras in learning research. In V. R. Lee (Ed.). *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 220–237). New York, NY: Routledge.
Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133.
Xie, K., Heddy, B. C., & Greene, B. A. (2019). Affordances of using mobile technology to support experience-sampling method in examining college students’ engagement. *Computers & Education, 128*, 183–198. https://doi.org/10.1016/j.compedu.2018.09.020.
Zollars, J. (2018). *Flow theory and engagement: Observing engagement through the lens of flow in a middle school integrated maker space.* Unpublished doctoral dissertationPittsburgh, PA: University of Pittsburgh.