Information and Learning Science

On researching activity tracking to support learning: a retrospective

Victor R. Lee,

On researching activity tracking to support learning: a retrospective

Victor R. Lee
Utah State University, Logan, Utah, USA

Received 13 June 2018
Revised 26 October 2018
26 November 2018
Accepted 29 November 2018

Abstract

Purpose – This paper aims to discuss research and design of learning activities involving activity tracking and wearable activity tracking technology.

Design/methodology/approach – Three studies are summarized as part of a program of research that sought to design new learning activities for classroom settings. The first used data from a qualitative interview study of adult athletes who self-track. The second used video excerpts from a designed learning activity with a group of fifth grade elementary students. The third study draws largely on quantitative assessment data from an activity tracking unit enactment in a rural sixth grade class.

Findings – Activity tracking appears to provide opportunities for establishing benchmarks and calibration opportunities related to intensity of physical activities. Those features of activity tracking can be leveraged to develop learning activities where elementary students discover features of data and how data are affected by different distributions. Students can show significant improvement related to statistical reasoning in classroom instructional units that centralize the use of self-tracked data.

Originality/value – As activity tracking is becoming a more ubiquitous practice with increased pervasiveness and familiarity with mobile and wearable technologies, this paper demonstrates a topical intersection between the information and learning sciences, illustrates how self-tracking can be recruited for instructional settings, and it discusses concerns that have emerged in the past several years as the technology related to activity tracking begins to be used for educational purposes.

Keywords Quantified self, Activity tracking, Elementary statistics, Personal informatics, Self-tracking, Wearables

Paper type Research paper

Introduction

It has become more common to see individuals of certain means sporting wearable devices that assist in their personal tracking of their physical activity. Often, these devices take the form of wristbands or smartwatches that track, at a minimum, the number of steps taken throughout the day and the calories expended. More expensive models will include features such as heart rate monitoring, sleep detection, amount of time spent sitting versus standing, number of floors equivalent climbed, minutes spent exercising, specific workouts completed

Members of the research team who contributed substantially to the work described here include Joel Drake, Ryan Cain, Jeffrey Thayne, Mary Briggs, Kylie Williamson, Michelle Berry, Ralph Trumble, Brad Buccambuso, Jon Thomas, Ani Aghababyan and Scott Smith. The author thanks Rebecca Reynolds and anonymous reviewers for feedback on earlier drafts. The author is grateful to teachers, students and participants in the various studies described. The work presented in this article was supported by a grant from the National Science Foundation under Grant No. DRL-1054280. The opinions expressed herein are those of the author and do not necessarily reflect those of the National Science Foundation.

Information and Learning Science
© Emerald Publishing Limited
2398-5348
DOI 10.1108/ILS-06-2018-0048

and locations traversed by way of GPS technology. The primary market for these technologies has been adults with disposable income who were seeking to pursue active lifestyles and maintain a level of ongoing fitness and health in pursuit of wellness.

Several years ago, just prior to activity tracking wearables becoming commonplace, my research group had begun a program of educational inquiry and design that examined the potential of these technologies to be sources of data that youth could collect and examine to support learning with and about data (Lee and Dumont, 2010). The underlying assumption had been that youth would be more knowledgeable about the data because they had produced it and had direct firsthand understanding (Hug and McNeill, 2008) of the activities that were represented by the data. Our first target population to help us examine those assumptions had been high school-age youth who used chest-worn heart rate monitors (built by Garmin) that communicated with wristwatches to log beats per minute based on electrical conductivity. At the time, the data transfer process was more onerous than what is required today, with today's technology typically involving a Bluetooth data transfer to a smartphone that happens regularly in the background of other activities. Back when that work began, obtaining records of electrical impulses from one's torso to infer heart rate was already cutting edge for a commercial wearable technology. To enable pilot students to examine the data, a specialized synching cradle that connected to a desktop or laptop computer had to be used. The computer that received the data needed specialized proprietary software for the data to be stored and viewed in any meaningful form. Now, those data are stored in the cloud. The data are viewed in custom apps with encouraging messages and prompts to be more active and earn badges for meeting goals that are configured into each device. Everyday conversation around activity tracking has shifted from "what is that thing around your wrist?" to "have you met your step goal today?" and "do you like that model better than the rival activity tracking technology?".

Regardless, my research team had different considerations in mind when beginning work with activity tracking technologies. Rather than focus on whether people were meeting fitness goals and whether the technology was effective at promoting wellness, we were curious about the questions that youth would pose if given access to such technology and how data about one's own activity could facilitate ways of thinking that were valued in mathematics and science. Indeed, we found through our earliest forays that students who were obtaining data from chest-worn heart-rate wearables began to interrogate the data based on what they already knew about themselves and physical activity. At the time, these technologies were used predominantly by athletes and fitness enthusiasts as personal training tools. In our original case study that represented our first published work on the topic (Lee and DuMont, 2010), we documented how a pair of teen girls who were asked to determine what activity was making them work harder, according to heart rate data, showed evidence of new learning about how to examine data. The pair that we studied had collected data from some common outdoor play activities (i.e. throwing a Frisbee and playing the basketball game "HORSE") and started their investigation by focusing on maximum values obtained in a single minute. Whoever had the highest single value was seen as having done harder work. However, they shifted to looking at where there were greater density of data points. Basically, they transitioned to considering more aggregate properties of the distribution of data points instead of making judgments based on a specific data point. What enabled this shift was personal views of who was more or less athletic and how simply looking at maximum values conflicted with their views of themselves. Also, their recall of the specific activities contributed, with students talking about whether they remembered had to do more or less moving during each recorded activity or not.

Since that original study, the devices and the broader discourses around the technology continued on their advance sensing the body, altering our views of self and health, and hacking wellness which is commonplace today (Lupton, 2014; Nafus, 2016). The primary market became all adults who felt they could afford to buy a device rather than athletes seeking to optimize their training and performance. The devices were furthermore biased toward specific audiences who were assumed to have the latest and greatest devices in a broader technology ecosystem. Still, as activity tracking devices became more familiar and cheaper to purchase, the devices became easier for the normatively able-bodied to wear and use. Our team adapted and migrated to new commercial devices for our research. The work extended to classrooms, with a focus on fifth and sixth grade students, as those grades were subject to educational standards expected proficiency in understanding how sets of data were organized and represented. Furthermore, those students were expected to be able to characterize sets of data by measures of center and spread. Those became research emphasis areas for several years, and in some ways continue to this day among members of my team (Drake, 2018; Thayne, 2016).

This article serves as an early retrospective, coupled with the presentation of some new data, on the use of commercial activity tracking technologies to support the design and implementation of new statistical learning activities for youth. As this is part of the launch of a new journal, this topic appeared to be an appropriate one as it spans concerns in both the information and learning sciences. Current research in information sciences takes seriously that the new forms of technology and information capture and storage around us are raising new questions for how to manage and act upon that information. Wearable technology represents one way in which information about physical activities is becoming a part of that landscape. Similarly, research in the learning sciences examines new technologies and arrangements for learning whether it takes place in formal settings (i.e. schools) or informal ones (i.e. homes, afterschool programs, museums, hobby groups, etc.). Wearable technologies have increasingly been an interest area in the learning sciences and have included, but also gone beyond activity tracking technologies (Lee and Shapiro, 2019). It seems apt for the inaugural issue of this journal to describe one line of work that has had some intersections with both areas.

In the sections that follow, I provide a brief summary of some relevant research that has been done in both information and learning sciences that serves to demonstrate that studying and designing for tracked activity data is one area of mutual interest. Following that, I provide summaries of three studies. One involves an examination of adult endurance athletes and how they collected and used their activity data. The information from that study informed a small design research study which sought to incorporate some of those findings into an activity we have called "quantified recess". That second design-based research study illustrates how the examination of one's own activity data can be a meaningful pursuit for elementary school students and can support increased understanding of outlier sensitivity in different measures of center. The third and final study is one from a classroom level design experiment that has not yet been reported. In that study, a full classroom unit was designed and implemented where each student in the class was provided with their own activity tracker to use during the school day. Through data obtained before and after that unit, I show how students in that class showed significant improvement in multiple areas on a validated learning progressions assessment instrument. Following that, I discuss what opportunities reside ahead for the use of tracked activity data for learning and what risks are associated. I also broaden the discussion to other related topics that we could explore through publication in this journal and through the continued dialogue between the information and learning sciences.

Activity tracking in information and learning sciences

Within information sciences, a modest but growing community has emerged that concerns itself with activity tracking and has often gone under the umbrella of "personal informatics" systems (Li *et al.*, 2010). Personal informatics can include a number of information types that can be collected relevant to a given individual that extends beyond physical activity. For instance, research and development projects in personal informatics that do not emphasize physical activity include those that help users examine their own productivity data and time spent on social networking sites by way of a workstation monitoring app (Collins *et al.*, 2014; Whittaker *et al.*, 2016). As personal informatics has grown as an interest area in the information sciences, and particularly in the field of human-computer interaction, frameworks and models for personal informatics have undergone further examination and refinement (Epstein *et al.*, 2015; Li *et al.*, 2011). Those modifications sought to explain better when and how individuals participate in technology-supported self-tracking. Other names used to describe this increased availability of self-tracked data include the Quantified Self, Personal Analytics, and Lifelogging (Lupton, 2014; Nafus, 2016; Lee, 2018). A tendency in these lines of research and development is to promote behavior change, with an underlying assumption that knowing more about our behaviors can empower users to then modify them. However, it is still unclear that behavior change is the dominant goal or that people who express an initial interest in self-tracking will maintain that commitment for an extended period of time (Clawson *et al.*, 2015).

Some projects in this area have examined archival data to see how participants in Quantified Self "meet-up" groups collect and share data (Choe *et al.*, 2014; Lee, 2014). Others have involved the design of technologies to provide new forms of feedback, such as a growing plant, in response to one's level of physical activity (Consolvo *et al.*, 2008) or design of new ways of representing activity such as showing a runner's current pace through LEDs on her shirt (Mauriello *et al.*, 2014). Interview and autoethnographic studies in human-computer interaction have also examined what influences and drives individuals to use various tracking technologies (Patel and O'Kane, 2015; Tholander and Nylander, 2015; Yang *et al.*, 2015). Together, these help to provide a more nuanced image of what compels people to use different tracking technologies and what new feedback systems are in development.

In the field of learning sciences, where design and study of new sociotechnical arrangements for learning are central concerns, there has been growing interest in mobilizing data about the self in service of learning. My own work has been situated as forms of quantified self in education (Lee, 2013a) and personal analytics for youth (Lee, 2018). However, I am not the only learning scientist who has examined self-tracking data experiences in educational settings. In museum settings, Lyons (2015) has used data that patrons produce as a strategy to promote understanding of increased challenges posed by climate change. In afterschool programs, Ching *et al.* (2016) have designed gaming activities to promote increased awareness of physically active lifestyles and also uncovered important constraints that youth face in making their own mobility decisions. Recently, Sommer and Polman (2018) described a learning sciences design project where a teacher worked with students in an alternative school setting to use Quantified Self data to develop infographics as a project to motivate student inspection and expression with data. Finally, in what would appear to be an overlap between information and learning sciences, Kang *et al.* (2016) have designed clothing for early elementary school aged children that embeds heart rate sensors and enables dynamic personalized data visualization and improved learning of human anatomy. In that project, youth learned more about anatomy and the impact of exercise on heart rate visualizations. Interest in quantified self in learning sciences even led to a keynote exploring overlaps between the two at the *2018 International Conference of the Learning*

Sciences and considerations the field should make as work in these areas moves forward (Tabak, 2018).

Taken together, there have been signs of mutual interest in both information and learning sciences. The core of activity-tracking research in both information and learning sciences is that some process or system is used to collect moderate to large amounts of data (on the scale of dozens to thousands of data points) generated by an individual through their behaviors. That individual then inspects those data in some manner so that they can generate new understandings. This allows for self-tracked data to enable a form of reflection that builds upon personal informatics systems (Baumer, 2015). Where exist some differences across information and learning sciences is in some of the methods and populations used. For instance, learning sciences tends to emphasize research on learners in designed learning settings. Information sciences goes into sociotechnical systems that are organized around a given form of work. Granted, high quality exceptions exist in both fields. However, there are ways in which those two directions of inquiry can be mutually informing. The next two studies are intended to illustrate one way that one line of work can greatly inform the other.

Study 1. Understanding the data practices and data relationships among endurance athletes

One strategy for understanding how people work with information, and in this line of work, self-tracked information, is to go out and understand what people do "in the wild". A qualitative study that I completed with my former doctoral student, Joel Drake, sought to do that. We interviewed 20 endurance athletes (runners and cyclists) recruited through running and cycling groups who self-identified as self-trackers. These interviewees spent time with interviewers showing the devices and records that they had created, explained why they collected those data, and how those data were useful or frustrating for them. In addition, they answered questions about the structure of bicycles (Lee, 2013b) and analyzed sample data obtained from tracking devices that did not belong to them. Details about these interviews, including the questions that were asked along with the coding processes, sample data excerpts, and some specific cases, is summarized in Lee and Drake (2013a).

One common observation was that several interviewees created their own logging systems. An example of such a logging system, not previously published but part of the original data corpus, appears in Figure 1. That particular log recorded number of minutes spent on each logged day doing designated exercise activities that ranged from shoveling snow, swimming, running, stretching, and hiking. That individual kept track of the number of hours each week she had exercised and had specific goals set. Her primary motivation was to avoid the difficulties that she had seen family members have as a result of histories of obesity. That participants' goal was to set goals and ensure accountability for them.

However, what we also discovered was that logged data took on many different roles for those who tracked. For some, logging served to provide benchmarks on previous performance. Knowing how one had performed on a particular exercise or what pace they maintained on the same distance of run in a previous year allowed for comparison and a form of competition with prior self. If previously obtained data were examined after new data were collected, then it was typically seen as feedback on how that athlete was doing relative to what they knew they could do. However, if previously obtained data were examined *before* collecting and examining new data, it served as a target to beat. The athletes wanted to improve over their past selves and had that in mind during their exercise or competition. This form of competition was seen as more considerate of where each athlete was in their performance and progress rather than positioning them in relation to elite athletes. (Elite benchmarks, such as what would qualify for the Boston Marathon, for

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Week Hours	Total Hours	Goal Hours	Week Miles	Total Miles	Goal Miles
Jan 1					Shvl Snow 15 Elliptical 60 (4) Weights 30	Running 105 (9.25)	Running 105 (9.25)	5.25	5.25	4	22.5	22.5	11 17
Jan 4	Spinning 60 (16) Swimming 90	Weights 30	Spinning 45 (12)			Running 60 (6) Elliptical 60 (4.25) Stretching 45	Walking 90 (6.5) Stretching 30	8.5	13.75	13	16.75 28	39.25 28	36 55
Jan 11	Spinning 60 (16)	Running 60 (5.75) Weights 45 1* Stretching 15 Swim 75	Spinning 60 (16)	Running 45 (3.5) Pyro 15	Swim 60 MC2 45 2* Dodge ball 45 Stretch 15 Weights 30 1*	Spinning 60 (16) Hiking 120 (6.25) Dry Canyon Stretch 30	Running 90 (7.5) Stretching 30	14	27.75	22	23 48	62.25 76	61 94
Jan 18	Spinning 60 (16) Stretch 15	Running 75 (6.5) Swim 90	Spinning 75 (20) Stretching 30 Weights 60 2*	Running 60 (5.25)	Shvl Snow 30	Spinning 60 (16) Running 90 (9) Shvl Snow 15		11	38.75	31	20.75 52	83 138	86 132
Jan 25	Spinning 60 (16)	Running 60 (6) Swimming 90	Spinning 60 (16)	Running 45 (3)		Spinning 60 (16) Running 75 (7.25)	Shvl Snow 45	8.25	47	40	16.25 48	99.25 186	111 171
Feb 1	Spinning 60 (16)	Running 45 (3) Weights 45 2* Swimming 75	Spinning 60 (16)	Running 60 (6)	Swim 60 TRX 45 2*	Skating 30		8	55	49	9 32	108.25 218	136 209
Feb 8	Spinning 45 (12) Weights 15 .5*	Running 60 (5.5) Weights 45 2* Swimming 75	Spinning 45 (12) Weights 15 .5*	Running 60 (5.25)		Running 120 (11) Stretch 30	Running 90 (8.5) Stretch 30	10.5	65.5	58	29.25 24	137.5 242	161 248
Feb 15	Spinning 60 (16) Stretch 15	Running 60 (5.5)	Weights 15 Spinning 45 (12) Stretch 15	Running 60 (4.5)	Swimming 75		Running 105 (10)	7.5	73	67	20 28	157.5 270	186 286
Feb 22	Spinning 60 (16) Weights 15	Running 60 (5.5) Weights 30 Swimming 75	Spinning 60 (16) Weights 15			Running 120 (10) Stretch 30		7.75	80.75	76	15.5 32	173 302	211 325

Figure 1.
A data log created and maintained by an adult endurance athlete that contains records of her exercise activities

instance, were still considered as those were valuable in their athletic communities even if they had no intention of participating in the Boston Marathon.)

Also, when and how data were presented raised different concerns. In Lee and Drake (2013a), a concern voiced by some was that having immediate data feedback pulled their attention from their immediate activity. That led them to "hide" the devices that were reporting their data (such as sticking it in their bags or turning them the other way so their screens were not visible) so they could look at the information after but not be absorbed by it during their exercise, event, or training. One very experienced participant, a woman who used a power meter along with a bike computer as part of her aggressive cycling training and had trained others, reflected this sentiment in the following statement:

> Participant: So, I compete in road biking and cyclocross, but my love is probably mountain biking. But I don't compete in mountain biking because I don't want to ruin it. I wanna keep it fun too [. . .]. I don't mountain bike with watts [tracked power] because I like it to be fun. I'm not out training. Not that I don't go hard [while mountain biking]!

For this participant, data (in the form of watts) detracted from the fun of mountain biking. That was the activity she loved and to compete and track data would "ruin it". It was not simply some mountain bike rides were ones that she did not want to track. Rather, that was her opportunity to still be a cyclist but have that completely separated from any sense of training that came with self-tracking.

This is consistent with other studies and writings on self-tracking. For instance, Buongiorno (2017) observed that the quantification of activities and explicit focus on just those quantities ran the risk of creating a detachment from the lived experience that produced those numbers. To avoid tracking would be to commit more fully to attending to the lived experience. Also, there was often an audience and judgment placed on what the numbers showed, according to an ethnographic study by Dudhwala (2017). In that study, when given a visualization of self-tracked data, there was an observed compulsion among some trackers to work harder so that they could make their data "look" better should others see the logged information. Excluding something from tracking prevents that judgment from being possible.

Another finding from this study was how different quantities took on situated meanings. For cyclists, a desirable cadence was above 80 rounds per minute. Using a cycling computer and sensors allowed cyclists to develop their intuitions for how 80 rpms felt in bodily effort. As they became increasingly familiar, several cyclists in that study reported not needing the cycling computer to tell them what was 80 rpms as they could judge that on their own. However, they did still use their cycling computers to track their rides and obtain overall statistics. Another example, which appeared in Lee and Drake (2013a), involved one individual who found that heart rate monitoring changed her outlook on her performance. As she described it, she had an expectation of being able to run at a certain level on inclined terrain but found she was unable to meet that expectation. As a result, she developed a critical view of her own abilities and considered herself to just be a "wimp" on inclines. However, when she had obtained a heart rate monitor and used it on inclined terrain, she discovered that her heart rate was over 200 beats per minute, meaning it was at her aerobic maximum. Her inability to keep her desired pace on inclines was not simply a personal failing. Rather, it was because her body was performing at its maximum level. To quote her, "[. . .] all of a sudden, it was 'I'm not a wimp. I'm pretty tough' if I can get it [my heart rate] to that high". The ability to quantify her performance and see those quantifications against some known benchmarks changed her view of herself and her performance in ways that were healthier and more productive for her eventual improvement.

Overall, findings from this study of activity tracking among adult athletes informed the design of an activity we developed with elementary school students. Among the key observations was to focus on relative change for individuals as performance targets for self-tracked activity so as to avoid a sense of competing against an unattainable elite benchmark. Another was to keep self-tracking narrow in scope so that a long term fixation on how their data looked or how important those numbers were could be diminished. We also came to appreciate how quantification could support a better sense of calibration for what numbers would be associated with different activities and how that tied into individual limits. The subsequently designed activity and how elementary school students responded to it are described in the next section.

Study 2. The quantified recess project

Drawing on lessons learned from the study of adult athletes who self-tracked with technology and also from the observation that seeing central tendencies and distributions were a potential for youth using activity monitor data, we developed an activity that we have called "quantified recess". This was a small scale effort on the escalating program of design-based research we were pursuing. This activity is described in more detail in Lee and Drake (2013b), along with a case study of how a fifth grade student and her assigned partner discovered outlier effects on different measures of center (means vs medians). That other paper describes how screen and webcam recordings were used for data collection to examine what youth were discussing in their data, what manipulations they had tried to change their data over time, what technology was used to convert data so that it was usable for students, and how that initial case was developed and analyzed to develop a narrative account. Briefly, the quantified recess activity was designed to encourage students to seek modifications in their physical activity during recess over the course of a single week. The technology used was a waist-worn Fitbit device (the Fitbit One waist-worn device) that recorded the number of steps taken each minute, which we provided back to students in a student-friendly data visualization software tool. Students were paired together with one student's performance being determined by the change in their mean steps per minute over the course of a week. The other student in the partnership was measured by the change in their median steps per minute over the course of the week. In total, 12 students (six teams) participated in this activity over two weeks (6 students during one week, then the other 6 during the next week).

The students who participated in the study revealed a number of intuitions that they tested about their recess movement that they could test and evaluate. Consistent with other research on personal intuitions about the operation of activity trackers, they did not always reflect on how the tracking devices actually worked (Yang *et al.*, 2015). For example, shuffle steps were thought of as a possible way to increase the number of steps taken, but then there was disagreement among students about whether those movements should count as a step. They were not registered by the wearable activity tracking devices as steps, which settled that matter for the students. Another question from students was if giving piggy-back rides to friends during recess could increase steps as one had to work much harder with the added weight of a friend to carry. They tried that as a way to increase their activity, but saw that did not increase the number of steps per minute in a way that was consequential for the metrics they were collecting. In the case study we presented in Lee and Drake (2013b), we provided a detailed account of a pair of students who were discussing how to accommodate for one student whose level of physical fitness did not allow her to be actively moving as fast as she could continuously for the entire recess. The score she was producing was based on the median number of steps per minute she took during recess. Like the adult athlete who used heart rate monitoring in the adult study described above, this student was finding that she was reaching her limit. She could not maintain a recess with nonstop continuous, high-intensity activity, while her assigned partner could. When the two students disagreed on what would be the best strategy to increase her score, she observed that breaks could be okay because those would be outlier values in her recess data. They would not influence the median level of activity if she stayed otherwise consistently during the times she was actively moving. That became their strategy to change their measured activity levels so that they were exceeding what they had done before. Their plan was to provide breaks for the one student who was struggling to increase her median score, knowing that it would not affect the new median she was trying to establish. Her partner, on the other hand, was going to stay as active as possible for as long as possible as his recess data were being represented by a mean and would be adversely affected by a break.

This became information that the entire group of students, beyond the focal pair, explored as the compared their strategies to manipulate their recess data over the week. The original case study pair shared what they had discovered on the last day when all students were examining their recess data together, and a different student's data visualization was used to explore possible manipulations to see if, as the case study pair had professed, that having outlier values because of taking breaks would be acceptable when considering median values. New for this paper is a brief account of what happened where that previous case study left off (not included in the original Quantified Recess paper due to space restrictions).

Beyond the pair of students in the previously published case, other students in the larger group who participated in quantified recess appeared to have gained an understanding of outlier sensitivity. This is illustrated with the brief transcript excerpt below, involving three students who were not in the originally profiled pair. The students, referred to with pseudonyms, were Callie, Anya, and Michael. Just prior to the moments indicated in the transcript excerpts below, several data points in Callie's data were changed (see Figure 2) by reducing their values to show a hypothetical a "break" in recess activity. This was done to verify that what had been shared before about outlier sensitivity. The activity facilitator asked why Callie's score did not change in response to those changes. Although they did not use formal statistical language, Anya explained her intuitions for why the median value did

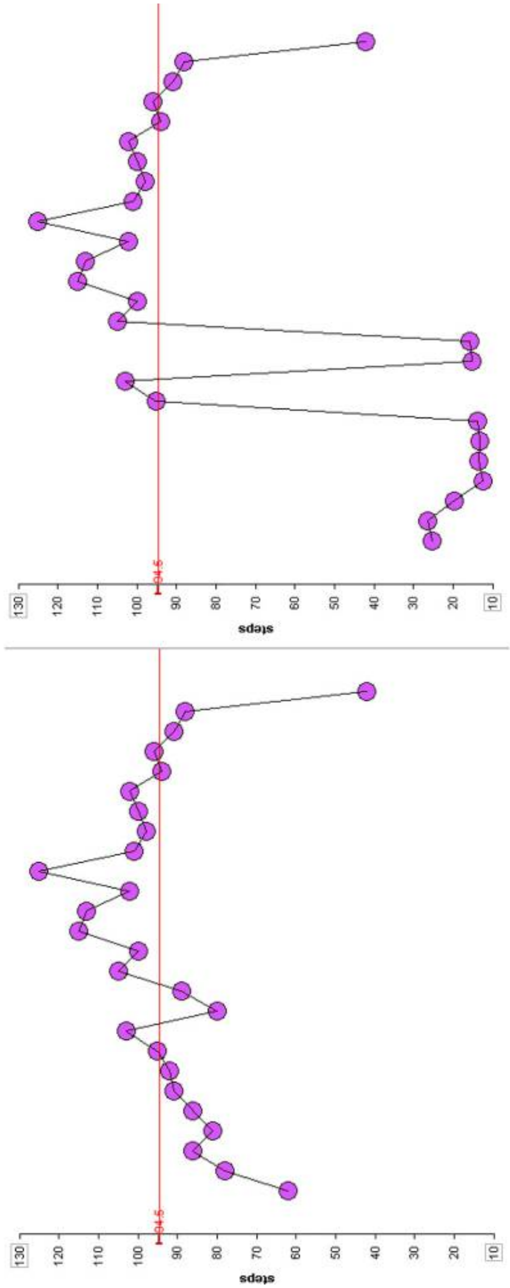

Figure 2.
Callie's recess data in steps taken per minute and the median value (left) followed by a manipulation where some values were artificially lowered and the median remains the same (right)

not change. Michael then restated the implications of it for how it would impact her recess score in the quantified recess activity.

Facilitator: Why isn't this [the change in values] changing Callie's score [median]?

Anya: When we moved those [data points], the ones close to the bottom, it [the median] didn't change – there was a lot, a lot more on the right side, up high.

Michael: Look, if she has 10 [minutes of] breaks and she's going hard for maybe 15 [minutes], it [the 15 minutes of hard activity] brings it [the score] up a lot. But it brings it up. If we had it at zero a lot, but if you go hard for like 14 and rest for like 11, it [the median] still would be higher.

In this exchange, Anya had referenced changing several data points that were below the median (red) line in the plot (Figure 2) and reiterating that the median value did not change. The reason for this was that there were more on the right-hand side of the plot that had high values relative to the lowered ones. The median served as a divider that split points across high and low values; it did not get affected from lower values being made even lower. When Michael spoke, he made a rough estimate of 10 break minutes being shown in Figure 2 by the drop of the data points. Those were akin to moving very little or not moving at all ("if we had it at zero a lot"). Students across settings referred to those as being time when a student was not actively moving about or taking "breaks". On the other hand, "going hard", whether it was for 15 or 14 min (as he offered both as suggestions), still meant there would be more data points on the higher side of the median line. While Michael may not have describe the median as splitting the data ordinally in half, he did recognize that even with changing some values to "breaks", the act of "going hard" would compensate for it. The difference in how Anya and Michael described changes in Callie's data was largely in how Anya focused on positions and values for data points, whereas Michael had translated those data points into recess action. Both represent ways of understanding medians and outlier sensitivity that go beyond what has been documented in other research with students across grade levels and adults (Jacobbe, 2008; Watson and Moritz, 2000). That Michael connected it to recess activity also suggested that thinking about this in terms of recess activity, rather than as data points, allowed that activity to serve as a meaningful context for this statistical reasoning.

While we did not perform systematic pre- and post-test measures of this group, we were seeing that students who were taking versions of this approach of using their own activity tracking for learning about and reflecting upon statistics content were showing greater learning gains about measures of center than those who were following the traditional curriculum over the same amount of time (Lee and Thomas, 2011). We also were discovering that recess had some especially useful properties within school for data inspection. Namely, it had much more active movement than most of the school day, students were doing a wide range of activities that were influenced by a multitude of social factors, it was a time when students had expertise on the physical activities that exceeded those of their classroom teachers, and it allowed for public recollection and testimony of what happened so that individual students' accounts of what took place on a given recess could be verified and challenged as necessary (Lee *et al.*, 2016). Also, students could calibrate their sense of recess activities with the number of steps per minute after relatively brief exposure to an activity tracking device. With this in mind, our team proceeded to develop multiweek curriculum to use in schools that could address all state standards and build on students intuitions about school day physical activity. As is to be expected in educational design-based research, multiple iterations followed. The team designed a classroom unit, implemented it with some teachers and students, refined it over the remainder of that year, and repeated that process the following year. This was then repeated one additional time during a still later school year. That final school year, after three years of classroom iterations outside of pilot work and small group activity trials like “quantified recess”, is the source of data for the final study described below.

Study 3. Implementation of a designed activity tracking unit with full classrooms

The last study I report on is work that has not appeared elsewhere, with the exception of some classroom conversations that had been analyzed and published in Lee (2017). We were interested in seeing what conceptual progress in elementary statistics content could be made by an intact class of students using a unit with quantified self data. This final study involved a sixth grade class of students at a rural school that had Title I status. No students or staff at the school had prior experience working with our team on this project. The teacher who led the class was in her third year of classroom teaching and had previously taught early childhood education before getting her teaching credential. As the school had been organized, the students were given a grade-level assessment at the beginning of the year to determine their mathematics performance. This particular teacher, whom we refer to as “Ms Hayley”, was given the lower performing half of students in the 6th grade for her math class. The higher performing students did math with the other sixth grade math teacher in the school. At this particular school, which served grades K-8, four classes of students existed for grades fifth and sixth. Each teacher for that grade level had a mix of fifth and sixth grade students throughout the school day. Math was one subject area where students were separated by grade level. Otherwise, activities were done across both grade levels. By the time we began working with this school, standards had been changed to move some of the content we had previously done with fifth grade students to sixth grade.

We were not provided formal data on students’ prior mathematics performance, but we were told by Ms Hayley and the school that Ms Hayley’s class of students included several who struggled with multi-digit subtraction and some of their core number sense skills. Indeed, we observed some of these difficulties when observing them during scheduled class time. Using rulers and calculators was also unfamiliar to most students in the class, making some traditional measurement and computational activities seemingly difficult to run without substantial support. There were also a number of students who had some additional special education needs that led to a full-time aide spending math period with Ms Hayley’s class. Like most rural schools, the students were predominantly white. We did not survey for family information, although the surrounding area was heavily reliant on the agricultural industry. On occasion, students would mention their families’ involvement in the industry by mentioning specific farms or companies where their parents worked. There were 29 students enrolled in Ms Hayley’s math class, although due to regular circumstances, one or more students was absent from school on any given day.

Procedures

All students in Ms Hayley’s class and their parent or guardian provided signed consent to participate in the research study, which was done as part of scheduled school activities. Ms Hayley’s class was provided with a mathematics and statistics unit that was designed iteratively by our research team and with assistance from past partner teachers. The unit was intended to last four weeks for one period of class each day of the week. With pre-planned field trips, assemblies, and scheduled release days, the unit lasted a little over five weeks into the school year. It was the first math unit done at the school following initial assessment and sorting of the students into the separate math classes. Ms Hayley also added two additional lessons of her own design that asked students to identify what were statistical questions and could be answered through quantitative research.

Every student in the class was provided with their own Fitbit Flex wristband device that would obtain step data throughout the school day. These devices were linked to anonymous accounts that we had used for other research studies in the interest of student anonymity to cloud-based servers. Each day of the unit, data that were automatically transferred to Fitbit servers was retrieved and provided in one minute intervals in *TinkerPlots* data visualization software (Konold and Miller, 2005). At the beginning of math class, which took place in the morning, the teacher identified one or two students whose activity data would be examined together by the class. Those students' data were projected, and the student associated with the data would narrate what activities they thought were represented in their projected data. Others in the class would offer their impressions and comment on sections where they thought something different than what the student had described was being projected. For example, students would sometimes dispute what was the recess activity that was represented and contend that the presenting student must have mixed up morning and afternoon recess activities in their narration and say why the data suggested as much. As noted in the discussion above, recess became a focal time for students as they had the most to say about those time periods. Not surprisingly, much of the school day involved step values of zero steps per minute as they were usually seated at their desks. We had anticipated this based on our quantified recess work and from prior classroom enactments. By the second week of the unit, students would only examine recess data when reviewing their own data.

As they examined the recess data, they identified ways to describe typicality and difference in the data that they had produced. This led to students developing, based on prompts from the teacher, procedures for identifying modes (with the teacher asking where the data seemed to clump together) and medians (with the teacher asking where seemed to be the "middle" point of the data). A fair share model for determining the mean was introduced by Ms Hayley after it began to get mentioned by students as an approach to determine what was "average" for the data. Different ways of describing how the data differed were introduced by students and those that matched normative techniques and terms in the standards were highlighted by Ms Hayley. By the third week of the unit, students were then tasked with manipulating their recess activities for one day to produce higher or lower measures of center or create the greatest amount of variation in their data. These were then examined on subsequent days along with some canonical representations of distribution, such as box and whisker plots. In addition, students worked on inventing histograms as a way to organize and represent data. That process involved students creating posters to show both typicality and difference in data, comparing their creations, and then taking the most useful ideas that appeared in class and working on a refined data plot. From there, the teacher helped the class to see how that was like a histogram, and the histogram plot was used regularly.

In the final week, based on various comments and questions that students had generated from ongoing review of their recess data, they worked in small groups to pursue a data investigation of their own choosing that involved comparison across two groups. Comparison of similar data sets had been previously identified in the statistics education literature as one effective means for supporting students' developing of inferential skills (Watson and Moritz, 1998). An example comparison we have documented in some detail at a different school in a previous year was completed by a group of students who sought to compare student step counts when playing (American) football and soccer (Lee *et al.*, 2015). In the present school, some students pursued comparisons of specific activities, some for specific days of the week, and some looked at comparisons across gender. On the last day of the unit, students made brief presentations of their data and what they felt they could conclude from their comparisons. All unit days were video-recorded with two video cameras following recommendations for camera placement and data management described in Derry *et al.* (2010). Those video recordings served as qualitative data for discourse and interaction analyses of students and teacher participation (Lee, 2018, 2019).

Assessment

For the purposes of evaluating implementation in Ms Hayley's class, students were given written pre- and post-tests using an assessment instrument developed by Lehrer *et al.* (2014). This assessment had been developed explicitly to inform how students were performing along a statistical thinking learning progression. In learning sciences research, learning progressions represent potential trajectories for student development of practices and content understanding in a specific disciplinary area (Duncan and Hmelo-Silver, 2009). Through an iterative instrument validation process, Lehrer *et al.* (2014) identified constructs that were being measured through their written instrument. One construct was "Conceptions of Statistics" (CoS), and it pertained to students abilities to describe distributions and measures of center (such as means, medians, and modes) in data, both through computational procedures and through qualitative estimation. Questions addressing this would ask for specific measures of center but also would ask what would happen to a measure of center if a data point were removed. Another construct was "Informal Inference" (InI), and addressed students' abilities to look at a samples of varied data and make inferences about population differences. For example, this could take the form of comparing two plots showing the number of bowling pins knocked down across several frames for two bowlers and inferring from that who seemed to be better at knocking down pins. "Modeling Variability" (MoV) was another construct that measured students abilities to account for differences in variability in distributions based on how measurement was done. An example of this would be to predict the differences in distributional shape if a less and more precise measurement instrument were used to obtain data. These constructs were assessed as part of a test that had seven scorable problem sets with subcomponent questions. The problem sets had combinations of multiple choice questions and open response questions.

Scoring

Lehrer *et al.* (2014) had developed assessment rubrics for each item in the written tests that were provided to us. These rubrics identified specific assessed constructs (i.e. *Conceptions of Statistics*, *Modeling Variability*, *Informal Inference*), a stated performance for each identified level, and examples of student responses that matched each performance level. These addressed both multiple choice questions and open response questions. It is important to note that each rubric item included between three-eight performance levels (inclusive of half levels when a performance was characterized as in transition but distinct enough to merit its own row in the rubric). This resulted in a complex scoring system designed to accommodate many degrees of nuance in student responses. Categorical placement of a student response to a single level was a challenging endeavor. The original instrument had been used with multiple sites and grade levels, making such nuance important in the original rubrics for showing how progressions on each construct varied and tied into each other. For our purposes, the instrument was for unit assessment, so we did not expect the entire range of the instrument to be represented from a single class. Still, we did observe that student responses could fit between identified levels in ways that had not been documented. A summary of the constructs and low, intermediate, and high performances as applicable to this group of students is shown in Table I.

One coder scored the open responses on the pre- and post-tests using a training set from a previous year to calibrate their judgments consistent with what was provided in the rubrics and to be consistent with how students in another school had been scored (Lee *et al.*, 2016). As mentioned above, there were several known numerical threshold levels for each item. In addition, there were also known sub-threshold levels within the provided rubrics that reflected a transitional performance. For instance, a *Conceptions of Statistics (CoS)* item may have had originally intended numerical threshold performance levels of CoS2 and CoS3 identified. For a specific item, subthresholds such as CoS2a through CoS2d (where a and d denote the subthreshold) had also been identified in the rubrics. For numerical scoring, the threshold and subthresholds were converted into a decimal scale with the whole number value taken from the threshold level (in the example of CoS2a, the threshold level number would be 2) and the decimal portion mapping each subthreshold letter onto a 0.2 increment (so CoS2a would convert to 2.2 and CoS2d would convert to 2.8). The 0.2 increment was important as well because there were still some known transitional values between subthresholds, denoted by a "-". This would be considered a 0.1 deduction on the unmodified score (so that CoS3c- would convert to 3.5).

Reliability

To check reliability, five students' pre- and post-tests were randomly selected for scoring by a different analyst to determine reliability. Based on exact matches at the subthreshold and transitional performance level, the two scorers achieved 87.4 per cent agreement with each scorer blind to the other's coding. Allowing for deviation of one known threshold level (i.e, one point), the scores had 89.6 per cent agreement. Due to the complexity of the scoring rubrics and the sparse and inherently ambiguous nature of handwritten student responses, we deemed this an appropriate level of scorer reliability. These percentages reflect only agreement for the open response questions and would be higher still if multiple choice responses (which are immediately agreed upon) were included, but we excluded those as their inclusion could artificially inflate reliability.

Table I.
Highest, intermediate and lowest performance levels on constructs of interest for our statistics unit

Construct	Lower performance	Intermediate performance	Highest performance
Conceptions of Statistics (CoS)	CoS1(a): Use visual qualities of the data to summarize the distribution	CoS2(a): Calculate the statistic for central tendency	CoS3(d): Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components
Modeling Variability (MoV)	MoV1(a): Attribute variability to specific sources or causes	MoV2(a): Informally estimate the magnitude of variation due to one or more causes	MoV2(b): Describe how a process or change in the process affects variability
Informal Inference (InI)	InI1(a): Make a judgement or prediction according to personal experience or beliefs	InI3(b): Compare two distributions based on one value such as the maximum, mean, or the number of cases above a cut-point	InI5(a): Compare two distributions based on proportions within defined regions, such as above/below a cut-point or IQR

Results

Twenty-six students (N = 26) in Ms Hayley's class completed both pre- and post-assessments. For evaluation purposes, we only included students who were present on days when the assessments were scheduled to be administered – the day before the unit began and the day after it was completed. As is to be expected in classroom-based research, some students were absent or had to miss part of the assessment period for reasons outside of our control. Students who were administered the pre- and post-assessment on other days were not included.

As shown in Table II and Figure 3, Ms Hayley's class showed a statistically significant improvement from pre-test to post-test based on performance on each of the three aforementioned constructs. Their initial performance on the Conceptions of Statistics

	Pre-assessment		Post-assessment	
Construct	Mean	SD	Mean	SD
CoS	0.142	0.148	1.182	0.580
InI	0.571	0.651	2.248	1.175
MoV	0.585	0.495	1.249	0.660

Table II. Mean performance for Ms Hayley's class on the pre- and post-assessments

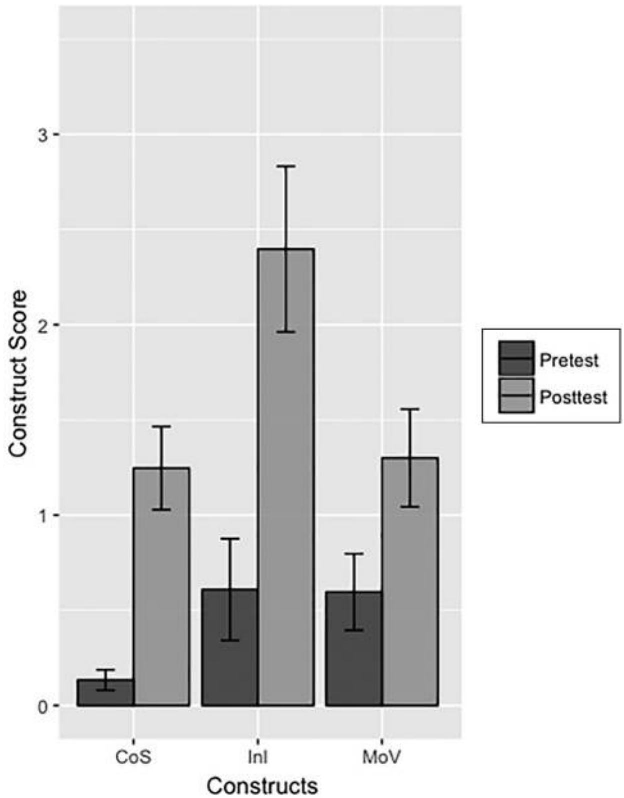

Figure 3. A plot showing pre- and post-assessment performance on the three constructs of Conceptions of Statistics, Informal Inference, and Modeling Variability

construct was 0.143 (sd = 0.148) but that improved to 1.182 (sd = 0.580) ($t = -10.32$, $df = 25$, $p < 0.001$, $d = 2.024$). Informal Inference performance was initially 0.571 (sd = 0.651) but that showed significant improvement to 2.248 (sd = 1.175) ($t = -7.105$, $df = 25$, $p < 0.001$, $d = 1.393$). Finally, students showed improvement in Modeling Variability, beginning at 0.585 (sd = 0.495) and finishing at 1.249 (sd = 0.660) ($t = -5.776$, $df = 25$, $p < 0.001$, $d = 1.133$).

Discussion of classroom enactment

Consistent with what we had accomplished in other classrooms (Lee and Thomas, 2011; Lee *et al.*, 2016), students showed significant improvement in their statistical understandings. What is notable about this class's result is it involved a different school than any that had been previously published and with a less experienced teacher (in number of years in the elementary school classroom) than in prior published studies. Additionally, it used another research team's validated assessment instrument to measure the performance of students. Thus, we find the results from this particular enactment encouraging. The approach of using activity tracking as a source of data for classroom instruction seemingly showed success.

Moreover, this was a class of students who were intentionally separated into a lower-performing math class but still showed significant improvement. We believe that this is because the data that students were using came from activities that they already knew. They were building new understandings from familiar activities that they were tracking, whether it was a game of tag during recess, using the jump rope, or playing four square. More details on some of the classroom interactions and finer-grained analysis of classroom discourse and conversations that support that assertion appear in other publications (Lee, 2018) or will be forthcoming in the future (Lee, 2019). Briefly, we have identified several instances of students referencing their knowledge of the specific activities and taking pride in being associated with their data that suggest students were more aware of what the data represented. For instance, we have seen students dispute whether 55 steps per minute on a data display would be closer to running or walking based on their own knowledge of what numbers they tended to record on their devices, students debating which projected set of data were representative of each student's recess because of how they thought data should have been shaped based on their knowledge of specific recess activities, and students explicitly recalling specific numerical values that they obtained when discussing overall class data distributions. Seeing students engage in these conversations suggests that they felt the numbers being used were consequential and relevant to their school day experiences. This had been the first time that Ms Hayley implemented the unit (and only time, as she left for a position at another school in a different city the following year), so we did not have a comparison available. However, in a recorded interview following the unit implementation, she did report to us that she felt very positively about the unit and the student experience. Especially noteworthy to her were occasions when students were deviating from planned step-tracking activities and animatedly identified their deviations on projected class data plots. According to Ms Hayley, seeing how their own behavior was quantified and related to others in the class was when she noticed students doing something very different related to elementary statistics instruction than what she had seen in her previous two years using her prior, school-issued curriculum which did not involve student-collected data. To her view, students working with their own tracked data that they could manipulate in their daily activities was what had made the difference as it made the statistical values more meaningful for them. This was a finding that was supported in a dissertation study that involved one of the project team members who helped develop the unit and collected data in Ms Hayley's class; in that study, the ability to intentionally manipulate quantities made those measured values more engaging for students (Thayne, 2016). A difference in that dissertation study was that the population involved was undergraduate students, rather than elementary students. However, the undergraduate students appeared to have a similar response in that the data they obtained of their own activities and used for analysis made the statistics content more comprehensible and compelling.

Summary and looking forward

This article reported primarily on three different studies that informed or refined the strategic use of activity tracking as a source of data for elementary students. The three studies, which are excerpted from a larger program of research that has and involved more youth, classrooms, and learning contexts than reported here, serve as one example of where information and learning sciences can have productive engagements with one another. Specifically, these projects touched upon practices that represent an important and growing area of research in information sciences such as personal informatics and self-quantification. One of the ways that this paper reflects that is through the first study of adult athletes. By looking at the data practices and meanings that motivated self-trackers used for their personal contexts, we gained some new understandings of how tracked data are meaningful and useful. Of special note was how competition with past performance was an especially meaningful metric, which informed the design of the quantified recess activity. Another was how self-quantification could serve to calibrate expectations and numerical values to physical activity. We found that inspection of quantifications of recess activity could enable learning of elementary statistics content. Specifically, we saw that students were recognizing what was reasonable for their bodies during intense recess activity and how outliers affect (or do not affect) measures of center.

These informed and led to larger unit designs that were scaled to entire classrooms where each student was provided with an activity tracking device and specific content goals tied to elementary statistics were targeted in the design of classroom lessons and whole class activities. Using an assessment instrument developed by another team, we found that significant improvement could be demonstrated even when students were classified as low achieving through a different assessment administered by their school and were seemingly in need of some remedial support. This is not to say that what we had engineered through design of a unit was the best possible outcome, but it was encouraging for us to accumulate more evidence that the approach we had been exploring could be made effective.

This particular article has been presented as a sort of retrospective on several years of research activity. As this article had opened, I had stated that the scope of self-tracking and its pervasiveness has changed substantially when my research group and I began to pursue this line of work. It is one that I expect we will continue to explore in the future, although much of the original novelty of the approach has become more routine and familiar. It has also begun to raise new questions that are worthy of further inquiry.

For example, and often posed in presentations of this work, are questions of privacy and consent. In the work we reported, and as part of university affiliated and federally funded research, informed consent has been a necessary and preferred mode of operation. Our precautions among the students and classrooms with which we have worked have included providing devices and attaching them to project accounts that cannot be matched to any of the participating students. Activity tracking that was used in the classrooms where we worked could not be traced back to any particular individual who was involved in our research. This is beyond any anonymizing and secure data storage methods we used as a matter of good research ethics. However, there is now increased awareness of how so many activities of ours and of children are being monitored by different tools and services. These are becoming a greater cause for concern. Imagine if, in several years, it becomes even easier for classes to work with data that students had obtained from their own activities. Say that a high school statistics class is able to support students in data inquiry using their own data. Perhaps the data that is used involves scraping mobile phones for songs most frequently played or lengths of text messages most often written or time spent on different apps. While data from such sources could have a great deal of personal relevance to students, it may quickly become information that they do not want to make public in the classroom or to their teachers. How do the potential benefits of building on data that are personally relevant conflict with potential concerns for privacy and risks for harm that may come from compromises in privacy? This is one tension that both information and learning scientists can jointly explore in the future.

Furthermore, as devices and data transfer have changed and become seemingly easier to use and initiate, there are open questions about tools that would be the most usable for students, teachers, and learners in a range of contexts. One promising example that touches upon expertise and interests in both information and learning sciences comes from the Scratch programming environment where Dasgupta and Hill (2017) have created sets of data coding blocks and data visualization tools that allow users to examine user behavior in the Scratch coding environment. In that case, the broad reach of Scratch as a widely used coding platform and the ability to inquire about what users do in Scratch opens up some new avenues for self-tracked data to be made into an educational resource.

Scratch has appeal in that it is free. However, another looming concern is equity of access to necessary tools and meeting the requirements for technological infrastructure. As it stands currently, automated activity tracking often uses separate wearable devices that are most compatible with the most current mobile devices, assume Wi-Fi access and presume some fluency with technology that is not universally shared. In one related project, a student and I worked with a different demographic than the one typically associated with self-tracking and discovered a number of infrastructural differences that affected full participation (Lee and Briggs, 2014). In that study, the youth involved did not necessarily have access to Wi-Fi or smartphones or models of smartphones that could retrieve and store data from various devices. Working with spreadsheets and new software also created additional overhead for students. Both learning scientists and information scientists must continue to contend with issues of inequitable access and implicit bias in the design of technologies and information experiences. Some institutions, such as schools and libraries, may provide common resources but are still being asked to address, on limited time and budgets, larger structural inequities that affect what information people can access and use. Indeed, these points about inequities and their consequences as they relate to self-tracking are cogently summarized from a sociological perspective by Lupton (2014). Moreover, activity tracking privileges individuals who are able-bodied. For full inclusion of youth given a range of mobility abilities, what options are there? In one year of our work, we had one student who was wheelchair-bound, and the accommodation was to use proxy data and for that student to concentrate on heart-rate data rather than step data. Still, an opportunity could have been pursued with rolling cadence on that student's wheelchair, or potentially GPS data to examine mobility given different movement options. It may have also been an opportunity to critically examine what spaces were and were not accessible. Representing all individuals and providing meaningful access remain core challenges that go beyond presumed technical infrastructure.

These challenges are mentioned to make more prominent the point that as this line of work began with my team, progress is being made, but obstacles still remain. New technologies have come to market that make some tasks easier to accomplish but they do not necessarily easier for all to access or reap benefits. With these changes and tensions, more collaborative work that merges expertise from different fields is needed. That work may look obviously like a study of learning, such as one that involves a classroom. It may also involve studies of different contexts where learning appears more subtly and is not the primary driver for participants (such as participating in a favored endurance event). In considering that this article is one of a handful that have the privilege of helping to inaugurate a new journal that brings information and learning sciences closer together and has sought to cover that range in a way that shows these lines of inquiry are mutually supportive, I am cautiously optimistic that the two fields can make some headway as more work is completed and reported by others in this journal in the future.

References

Baumer, E.P.S. (2015), "Reflective informatics: conceptual dimensions for designing technologies of reflection", Paper presented at the Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul.

Buongiorno, F. (2017), "A phenomenological understanding of digital processes of subjectification: the example of lifelogging", *Philosophy Study*, Vol. 7, pp. 341-349.

Ching, C.C., Stewart, M.K., Hagood, D.E. and Rashedi, R.N. (2016), "Representing and reconciling personal data and experience in a wearable technology gaming project", *IEEE Transactions on Learning Technologies*, Vol. 9 No. 4, pp. 342-353.

Choe, E.K., Lee, N.B., Lee, B., Pratt, W. and Kientz, J.A. (2014), "Understanding quantified-selfers' practices in collecting and exploring personal data", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Toronto, Ontario*.

Clawson, J., Pater, J.A., Miller, A.D., Mynatt, E.D. and Mamykina, L. (2015), "No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist", *UbiComp 2015, ACM, Osaka*.

Collins, E.I.M., Cox, A.L., Bird, J. and Harrison, D. (2014), "Social networking use and RescueTime: the issue of engagement", *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM, Seattle, Washington, DC*.

Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., Lamarca, A., Legrand, L., Libby, R., Smith, I. and Landay, J.A. (2008), "Activity sensing in the wild: a field trial of ubifit garden", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Florence*.

Dasgupta, S. and Hill, B.M. (2017), "Scratch community blocks: supporting children as data scientists", *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, Denver, CO*.

Derry, S.J., Pea, R.D., Barron, B., Engle, R.A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J.L., Sherin, M.G. and Sherin, B. (2010), "Conducting video research in the learning sciences: guidance on selection, analysis, technology, and ethics", *Journal of the Learning Sciences*, Vol. 19 No. 1, pp. 3-53.

Dudhwala, F. (2017), "Doing the self: an ethnographic analysis of the quantified self", PhD thesis, University of Oxford.

Duncan, R.G. and Hmelo-Silver, C.E. (2009), "Learning progressions: aligning curriculum, instruction, and assessment", *Journal of Research in Science Teaching*, Vol. 46 No. 6, pp. 606-609, doi: 10.1002/tea.20316.

Drake, J. (2018), *Riding to Learn: Informal Science in Adult Cycling Communities*, UT State University, Logan, UT.

Epstein, D.A., Ping, A., Fogarty, J. and Munson, S.A. (2015), "A lived informatics model of personal informatics", *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, Osaka*.

Hug, B. and McNeill, K.L. (2008), "Use of first-hand and second-hand data in science: does data type influence classroom conversations?", *International Journal of Science Education*, Vol. 30 No. 13, pp. 1725-1751.

Jacobbe, T. (2008), "Elementary school teachers' understanding of the mean and median", in Batanero, C., Burrill, G., Reading, C. and Rossman, A. (Eds), *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference, Monterey, Mexico*.

Kang, S., Norooz, L., Oguamanam, V., Plane, A.C., Clegg, T.L. and Froehlich, J.E. (2016), "SharedPhys: live physiological sensing, whole-body interaction, and large-screen visualizations to support shared inquiry experiences", *Proceedings of the The 15th International Conference on Interaction Design and Children, ACM, Manchester*.

Konold, C. and Miller, C. (2005), *TinkerPlots: Dynamic Data Exploration: Statistics Software for Middle School Curricula*, Key Curriculum Press, Emeryville, CA.

Lee, V.R. (2013a), "The quantified self (QS) movement and some emerging opportunities for the educational technology field", *Educational Technology*, Vol. 53, pp. 39-42.

Lee, V.R. (2013b), "Knowing and learning with technology (and on wheels!): an introduction to the special issue", *Technology, Knowledge and Learning*, Vol. 18 Nos 1/2, pp. 1-8.

Lee, V.R. (2014), "What's happening in the quantified self movement?", in Polman, J.L., Kyza, E.A., O'Neill, D.K., Tabak, I., Penuel, W.R., Jurow, A.S., O'Connor, K., Lee, T. and D'Amico, L. (Eds), *The International Conference of the Learning Sciences (ICLS) on Learning and Becoming in Practice*, Vol. 2, ISLS, Boulder, CO, pp. 1032-1036.

Lee, V.R. (2018), "Personal analytics explorations to support youth learning", in Zheng, R. (Ed.), *Digital Technologies and Instructional Design for Personalized Learning*, IGI Global, Hershey, PA.

Lee, V.R. (2019), "Supporting complex multimodal expression around representations of data: experience matters", in Sengupta, P., Kim, B. and Shanahan, M.-C. (Eds), *Critical, Transdisciplinary and Embodied Approaches in STEM Education*, Springer, Berlin.

Lee, V.R. and Briggs, M. (2014), "Lessons learned from an initial effort to bring a quantified self 'meetup' experience to a new demographic", *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM, Seattle, Washington, DC*.

Lee, V.R. and Drake, J. (2013a), "Digital physical activity data collection and use by endurance runners and distance cyclists", *Technology, Knowledge and Learning*, Vol. 18 Nos 1/2, pp. 39-63.

Lee, V.R. and Drake, J. (2013b), "Quantified recess: design of an activity for elementary students involving analyses of their own movement data", in Hourcade, J.P., Miller, E.A. and Egeland, A. (Eds), *Proceedings of the 12th International Conference on Interaction Design and Children 2013, ACM, New York, NY*.

Lee, V.R. and Dumont, M. (2010), "An exploration into how physical activity data-recording devices could be used in computer-supported data investigations", *International Journal of Computers for Mathematical Learning*, Vol. 15 No. 3, pp. 167-189.

Lee, V.R. and Shapiro, R.B. (2019), "A broad view of wearables as learning technologies: current and emerging applications", in Diaz, P., Ioannou, A., Bhagat, K.K. and Spector, J. M. (Eds), *Learning in a Digital World - Perspectives on Interactive Technologies for Formal and Informal Education*.

Lee, V.R. and Thomas, J.M. (2011), "Integrating physical activity data technologies into elementary school classrooms", *Educational Technology Research and Development*, Vol. 59 No. 6, pp. 865-884.

Lee, V.R., Drake, J.R. and Thayne, J.L. (2016), "Appropriating quantified self technologies to improve elementary statistical teaching and learning", *IEEE Transactions on Learning Technologies*, Vol. 9 No. 4, pp. 354-365.

Lee, V.R., Drake, J., Cain, R. and Thayne, J. (2015), "Opportunistic uses of the traditional school day through student examination of fitbit activity tracker data", in Bers, M.U. and Revelle, G. (Eds), *Proceedings of the 14th International Conference on Interaction Design and Children*, ACM, Boston, MA, pp. 209-218.

Lehrer, R., Kim, M.-J., Ayers, E. and Wilson, M. (2014), "Toward establishing a learning progression to support the development of statistical reasoning", in Confrey, J., Maloney, A. and Nguyen, K.H. (Eds), *Learning over Time: Learning Trajectories in Mathematics Education*, Information Age Publishers, Charlotte, NC.

Li, I., Dey, A. and Forlizzi, J. (2010), "A stage-based model of personal informatics systems", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM*, *Atlanta, GA*.

Li, I., Dey, A.K. and Forlizzi, J. (2011), "Understanding my data, myself: supporting self-reflection with ubicomp technologies", *Proceedings of the 13th International Conference on Ubiquitous Computing, ACM*, *Beijing*.

Lupton, D. (2014), "Self-tracking cultures: towards a sociology of personal informatics", *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design, ACM*, *Sydney, New South Wales*.

Lyons, L. (2015), "Exhibiting data: using body-as-interface designs to engage visitors with data visualizations", in Lee, V.R. (Ed.), *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*, Routledge, New York, NY.

Mauriello, M., Gubbels, M. and Froehlich, J.E. (2014), Social fabric fitness: the design and evaluation of wearable E-textile displays to support group running", Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Ontario.

Nafus, D. (2016), *Quantified: Biosensing Technologies in Everyday Life*, The MIT Press, Cambridge, MA.

Patel, M. and O'Kane, A.A. (2015), "Contextual influences on the use and non-use of digital technology while exercising at the gym", Paper presented at the Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul.

Sommer, S.R. and Polman, J.L. (2018), "From quantified self to building a more fit community: data tracking and science infographics as boundary objects", in Kay, J. and Luckin, R. (Eds), *13th International Conference of the Learning Sciences (ICLS) Rethinking Learning in the Digital Age: Making the Learning Sciences Count*, Vol. 2, ISLS, London, pp. 1001-1004.

Tabak, I. (2018), "Digital footprints and shoes that don't fit", in Kay, J. and Luckin, R. (Eds), *13th International Conference of the Learning Sciences on Rethinking Learning in the Digital Age: Making the Learning Sciences Count, (ICLS) 2018*, *ISLS*, *London*.

Thayne, J. (2016), *Making Statistics Matter: Using Self-data to Improve Statistics Learning*, UT State University, Logan, UT.

Tholander, J. and Nylander, S. (2015), "Snot, sweat, pain, mud, and snow: performance and experience in the use of sports watches", *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM*, *Seoul*.

Watson, J.M. and Moritz, J.B. (1998), "The beginning of statistical inference: comparing two data sets", *Educational Studies in Mathematics*, Vol. 37 No. 2, pp. 145-168.

Watson, J.M. and Moritz, J.B. (2000), "The longitudinal development of understanding of average", *Mathematical Thinking and Learning*, Vol. 2 Nos 1/2, pp. 11-50.

Whittaker, S., Kalnikaite, V., Hollis, V. and Guydish, A. (2016), "'Don't waste my time': use of time information improves focus", Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA.

Yang, R., Shin, E., Newman, M.W. and Ackerman, M.S. (2015), "When fitness trackers don't 'fit': end-user difficulties in the assessment of personal tracking device accuracy", *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM*, *Osaka*.

Further reading

Brown, A.L. (1992), "Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings", *Journal of the Learning Sciences*, Vol. 2 No. 2, pp. 141-178.

Rapp, A. and Tirabeni, L. (2018), "Personal informatics for sport: meaning, body, and social relations in amateur and elite athletes", *ACM Transactions on Computer-Human Interaction*, Vol. 25 No. 3, pp. 1-30.

About the author

Victor R. Lee is an Associate Professor of Instructional Technology and Learning Sciences at Utah State University. He obtained PhD in Learning Sciences from Northwestern University. Victor R. Lee can be contacted at: victor.lee@usu.edu