# Appropriating Quantified Self Technologies to Support Elementary Statistical Teaching and Learning

Victor R. Lee, Joel R. Drake, and Jeffrey L. Thayne

**Abstract**—Wearable activity tracking devices associated with the Quantified Self movement have potential benefit for educational settings because they produce authentic and granular data about activities and experiences already familiar to youth. This article explores how that potential could be realized through explicit acknowledgment of and response to tacit design assumptions about how such technologies will be used in practice and strategic design for use in a classroom. We argue that particular practical adaptations that we have identified serve to ensure that the classroom and educational use cases are appropriately considered. As an example of how those adaptations are realized in actual elementary classrooms, we describe an effort to provide fifth-grade students each with their own Fitbit activity trackers in the context of a multi-week unit exploring core ideas in elementary statistics. Observational descriptions and transcript excerpts of students and teachers discussing their own Fitbit data are presented to illustrate what opportunities exist to leverage youth familiarity with daily activities in a way that targets development of statistical thinking. Quantitative written test results showing learning gains and differences between traditional and wearable device-enhanced instruction are also presented. Improvement on several statistical thinking constructs is identified, including in the areas of data display, conceptions of statistics, modeling variability, and informal inference.

**Index Terms**—Computers and education, user generated learning content, devices for learning

✦

## 1 INTRODUCTION

RESEARCHERS in human-computer interaction and ubiquitous computing have been taking notice of what has been called the "Quantified Self" (QS) movement [1], [2], [3], [4], [5]. Much of this movement has been characterized by rapid growth and availability of both wearable and smart devices that obtain and store everyday data about specific individuals and their routine activities. For instance, wi-fi scales enable new ways to monitor body weight, and wrist-based activity trackers offer new ways to evaluate how active or sedentary our workdays have been. This has been an intriguing sociotechnical development because data about the mundane are now broadly available and made potentially intriguing. As a new class of mobile devices, QS wearables are enticing because they enable personal examination and interrogation of everyday engagements with the world [6].

At the same time that we see the rise of QS technologies, recent education standards have begun to push for students to develop more fluency with data and statistical reasoning. These are to take place starting in elementary school [7], [8]. Our position on this is that introduction of data-generating QS technologies into elementary classrooms may allow us to support those goals, but only if there is deliberate effort by researchers to seriously consider what such an introduction actually entails.

Defending that position is the overarching goal of this article. By way of demonstration case, we seek to provide some initial answers to the following questions: (1) How can wearable QS activity tracking devices be meaningfully integrated into the elementary classroom to support productive engagements with quantitative data? (2) To what extent do students demonstrate greater competency in statistical reasoning as a result of engaging in such experiences? To date, the nascent wearable technology literature involving teaching or learning focuses on out-of school settings [5], [9]. For classrooms, the field only has limited examples that either speak to previous generations of wearable technology [10], or are brief descriptive case studies of students working with wearable technology to support content learning [2], [11]. This study is intended to broaden and extend that literature base.

Part of our main argument here is that we should take seriously that there is a great potential for QS technologies to be used in classrooms. At the same time, we recognize that QS technologies have emerged in one use context where certain assumptions are implicitly made about the user and the surrounding technological ecosystem. The next section (Section 2) identifies some of those tacit assumptions, where there are obstacles for porting those, and how the obstacles can be addressed in the case of school use. Then we summarize some of the current literature related to how students often fail to learn to work meaningfully with data in elementary school classrooms. This illustrates why classrooms need new approaches. We then discuss what newly emerging pedagogical approaches and technological tools may be the most usefully combined with wearable QS technologies. That three-part focus on how QS and classrooms fundamentally differ, where current educational approaches have

● *The authors are with Utah State University, Logan, UT 84322.*
  *E-mail: victor.lee@usu.edu, {joel.drake, jeffrey.thayne}@aggiemail.usu.edu.*

fallen short, and what tools and teaching strategies provide the most promise together create the foundation us to report on observations and results from of a particular iteration of an educational "design-based research" project that brings QS wearables into the classroom.

For those less familiar, educational design-based research has some similarities to human-computer interaction design and research practices such as participatory design or contextual design [12], [13]. Direct engagement and observation with students and teachers help educational design researchers to identify and develop practical solutions related to the use of new technologies in the classroom (Section 3.1). The larger corpus of user engagements, which we had videorecorded, transcribed, and iteratively reviewed, also served as the source of observational report and transcript excerpts from Section 3.2. We present those to help convey what an elementary statistics unit with wearable QS technology could look like. Following that description, we describe a learning measurement instrument (Section 4), a comparative assessment approach (Section 5), and some comparisons of resultant student learning from a unit involving QS technologies (Section 6). Finally, we offer some reflection and discussion of what this effort implies for those more broadly interested in how wearable technologies could be used in the classroom (Sections 7 and 8).

## 2 DIFFERENCES, CONSTRAINTS, AND OPPORTUNITIES FOR QS IN THE CLASSROOM

Several key differences exist between power users who are part of the QS movement and the students in an elementary classroom. First, users involved in QS are usually very comfortable in knowing how to read and interpret data. In an elementary classroom, being able to read and interpret data is the *goal* rather than a prerequisite. For QS technologies to be usable for students, support has to be introduced in the form of deliberate lessons and sequenced activities that enable students to know what they are seeing when they see automatically captured device data. As a core tenet of Constructivism in education is that new knowledge must develop from old [14], the most fruitful approaches would go beyond a lecture on how to read data. Instead, lessons would involve students doing active sense-making with peer and teacher support. Because QS deliberately focuses on data about activities that are mundane and familiar to specific individuals, there is a unique opportunity to support this tenet of Constructivism. The mundane and familiar serves as old and pre-existing, but usable, knowledge. That develops into more sophisticated statistical knowledge later. This continuity represents a unique learning affordance of wearable devices compared to other mobile technologies.

Another difference between more users in the QS community and a classroom of elementary students is access to other support technologies. Many wearable and smart devices require that a user have another mobile device, a laptop, or at least a desktop computer so that data can be transferred and later accessed. Those are typically necessary to pair their wearable with another device. Those two personal computing devices can then repeatedly synchronize in the background while both devices remain on their person. Elementary students and elementary classrooms rarely have

access to these technologies. In a typical school classroom, which has students from a range of backgrounds and circumstances, it is not even a correct assumption that most students have home access to current computing systems or the high-speed connectivity necessary to fully use all wearable device functionalities [15]. For students to be able to access their data, some work must be done that enables elementary students of variable means to take advantage of what limited computing resources are already in the classroom.

Still another important distinction between QS and school communities is differences in daily schedules. For example, data obtained from a Fitbit device that would be used by a QS-er are reported back in 15-minute increments. This often suits an adult who has scheduled events that take place on the hour or on the half hour. However, a typical school day is partitioned very differently. Say that a student wanted to see how their in-between class activity compared to their in-class activity. If there is a passing period between classes from 2:05 to 2:10 but a Fitbit interval is fifteen minutes long, we cannot determine how many calories in that interval could be attributed to a five-minute period. This means that for data to be useful in school, they must be accessible in different interval sizes. That is a capability that requires some additional programming and database querying.

Finally, self-quantification among experienced adult users often relies on data collected over long stretches of time. A QS-er may have tracked and analyzed their daily weight over a period of two years to see if there were any seasonal changes or how often their weight goals were met. Neither a student nor a classroom teacher has the luxury of collecting that kind of longitudinal data. This makes the classroom capture of sufficient data points to see noteworthy trends difficult. However, one advantage for the classroom is the presence of dozens of students who are doing many of the same activities. By capturing and examining data that are more *lateral* than *longitudinal*, it could be possible for students to still leverage data that are theirs, get a large number of data points, and do so in the more constrained period of time that schools face. In that regard, for the purposes of the classroom, the activities may be better conceptualized at their onset as being about *Quantified Selves* rather than an individual *Quantified Self*.

### 2.1 How Children are Typically Exposed to Statistics in Elementary School

Beyond acknowledging differences in QS practice and in classrooms, there are other considerations one must make when bringing QS technologies to school. For instance, we should consider how schools currently teach students to use quantitative data and if there is even a need to innovate.

Currently, most elementary classrooms cover some statistical content, but it turns out the approaches most often represented in elementary textbooks use conceptual models that undermine student learning [16]. The typical content addressed in elementary school includes measures of center and the range of a set of data, often with a small number of values given (i.e., less than 10). Students are presented with the algorithmic procedures for calculating these values, with which they can eventually demonstrate mastery (or more likely, memorization). Some of this limited presentation may be due to teachers' own discomfort with the subject

matter [17]. As it stands, there is an extensive literature that demonstrate that direct recall and mechanical use of a known algorithm represents the most depth of students' understandings of fundamental statistical concepts, such as the mean [18]. That is, they can only recite a procedure but cannot explain why the procedure is a good one or why it works. As these students continue in the educational system, they then struggle in dealing with novel data-related problems, connecting statistical measures to the data they characterize, or making meaningful inferences from data [19]. Moreover, those difficulties persist into adulthood [20], [21].

In addition, learners often fail to develop robust undesrstandings of crucial properties for different measures of center, such as how they differ from each other, or in what contexts where each is the most useful [22]. One important example is the fact that the mean is sensitive to outlying values. A mean can, under certain circumstances, differ substantially from the median in seemingly counterintuitive ways. For instance, we might characterize household income as having an increasing mean but a decreasing mean, relative to decades past. As it stands right now, the literature suggests we are not currently enabling students to understand such distinctions or other core statistical ideas that will be consequential for their adult lives.

## 2.2   How Students Could Productively Learn Statistics Ideas

The previous section confirmed that at present, children have substantial conceptual difficulty with elementary statistics. Under what conditions do children show conceptual competence? That question is the concern of this section.

More optimistic strands of educational research suggests children can become quite competent with data and statistics if they are first made comfortable with the epistemic uncertainty inherent in measurement and data creation [19]. When scientists obtain measurements, they intentionally obtain enough measurements so they can identify regularities and tendencies within their data. They do that to confront the inherent uncertainty and unavoidable error that comes from measurement.

Prior design-based classroom research with elementary students has suggested error and uncertainty can and should be introduced fairly early to elementary school students [23], so that they can view data-related activities as an exercise in obtaining enough information to ascertain what, within a data set, is the "signal" and what is the "noise" [24]. Instructionally, this means effective statistics instruction should centralize variability as a big idea underlying data. In part, this is more consistent with the historical development of statistical inquiry; the historical argument that has been made in favor of variability as a core and early idea is that statistical measures and concepts actually arose largely because conceptual and computational tools were needed to explain and model variability in measurement [25]. The literature from science education, another area in which students learn to collect and make sense of data, also suggests that first-hand experience with data is profitable [26]. Students are better equipped to talk about properties in a set of data if they were involved in collecting those data than if they were provided with some data set that had been obtained without their involvement. In thinking about this

current study and how appropriation of quantified self technologies would be sensible, we reiterate that QS technologies are designed to exploit routine familiarity. By engaging in routine activities, students can be directly involved in data collection experiences, making them potentially promising as educational tools.

## 2.3   Visualization Technologies to Support Statistics Learning

Educational technologists have previously developed some tools to support visualization demands associated with students' statistical inquiry that leverage children's intuitions. Examples include agent-based modeling tools and participatory simulations where students play the role of individual data points or obtain several data samples from a larger population [27], [28], and computational minitools [29] that show distributions from hypothetical testing scenarios. In addition, there have been data visualization tools for children that depict inspectable data points that can be manipulated through direct interface controls, such as drag and drop. One of the earliest incarnations was the *Tabletop* software tool [30] that showed different groupings of data that a user could customize in order to performa exploratory data analysis. More recently, *TinkerPlots* [31], emerged as a next generation software tool intentionally designed and built to enable students in grades 3-8 to explore statistical data.

*TinkerPlots* was our visualization tool of choice for this project. This was in part because its development of *TinkerPlots* has been informed by several years of empirical research on how students in our target age range intuitively reason about data [32]. Like *Tabletop*, it employs the use of individual icons, which are by default a filled circle, as visual depictions of individual data records. A user can sort, organize, and color-code their data based on several different parameters. Transitions in data organization are shown through dynamic animations where the user sees all the data points moving to their respective places.

By default, *TinkerPlots* comes with several sample data sets, and users can also import new datasets using properly formatted .csv files. We leveraged this capability to import data obtained from wearable activity trackers. The benefit of using *TinkerPlots* is that, unlike the standard data dashboards provided for owners of wearable activity trackers [5], the interface of *TinkerPlots* is pedagogically designed so that only a few steps are needed in order to produce a density plot or a histogram. A version of this tool, known as *Fathom* [33], is available for higher grade levels, and an open-source version of both tools (CODAP) is currently under development with support from the National Science Foundation through the Concord Consortium.

## 3   IMPLEMENTATION

### 3.1   Technologies

#### 3.1.1   Fitbit Ultra and Fitbit One

For various practical reasons, including data accessibility through a partner API, we opted to use Fitbit Ultra and Fitbit One devices for our classroom work (Fig. 1). At their most basic, Fitbit devices are digitally enhanced pedometers. A traditional, inexpensive pedometer usually has some

Fig. 1. The Fitbit one (left) was one of the wearable fitness tracking devices used for this study. Activity data are displayed on the Fitbit dashboard (fitbit.com) using time-ordered and aggregate representations (right).

sort of very simple oscillation mechanism, such as a spring or small pendulum, in plastic casing. This oscillator moves while a person walks. Certain amplitudes in a given oscillation cycle will hit a switch or lever and increase the step counter by one, leading to step counts. Current wearable activity trackers, including those made by Fitbit, instead use a three-axis accelerometer to detect movement. Based on sensed changes in three-dimensional space and using a proprietary algorithm, the device categorizes movements as steps or general activity. This can be inferred based on specific signatures associated with step motions that go beyond a single spike in amplitude.

The current number of steps is shown on an OLED display and stored in one-minute increments. In addition to steps, these Fitbit devices can also capture data such as equivalent flights of stairs ascended (which are also inferred based on accelerometer data, although other tools can use barometric pressure data to infer elevation gain) and will compute estimates of activity level, distance walked, and calories expended based on motion data and user profile information. Because there is on-board memory with Fitbit devices, more granular data can be recorded over time than a simple pedometer that only stores a running total. Captured data are wirelessly uploaded to Fitbit servers periodically whenever the device is within range of a connected computer or smart device via Bluetooth. Of all the available data in Fitbit devices, we chose to focus students' attention on steps because steps were the easiest data for students to relate to and the closest to "raw."

As with any measurement device, there is variability in a Fitbit's step counting accuracy. Step motions can be feigned through some other activities (such as jumping) or unexpected jiggles that incidentally bear resemblance to a footstep movement signature. With that said, the device is programmed to avoid false readings when possible and can often avoid counting abrupt shaking (something that children often like to do with a step counting device) as steps. The general accuracy in laboratory conditions for Fitbit devices in terms of step counting is considered relatively high [34], which is leading it to be increasingly used in public health research.

### 3.1.2 A Custom Data Grabber

Fitbit activity trackers record data at 1-minute intervals, even though those are not displayed on the dashboard nor accessible to general users. Through the Fitbit Partner API, third-parties may access this "raw" data. Using this API, we developed a basic web form (http://ecds.ed.usu.edu/fitbit2) that allows us to request up to three weeks of minute-by-minute data at a time. These data (including steps taken, calories expended, and floors climbed) are returned in JSON format.
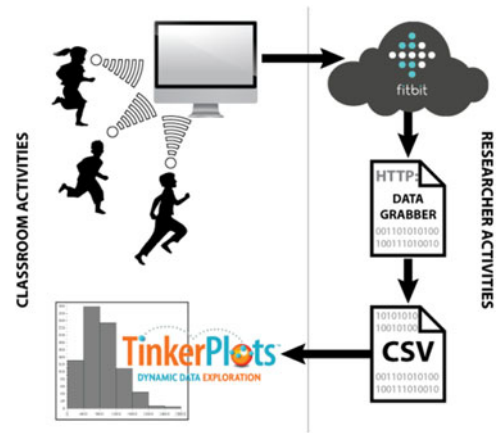


Fig. 2. Fitbit data collection process. Student activity data tracked by Fitbits wirelessly syncs to the Fitbit servers. Using the Data Grabber developed by the research team, CSV files are obtained. The CSV is imported into TinkerPlots for display and student analysis.

A custom script parses the returned values and outputs them to a browser window as comma separated values. The user can then save this output as a CSV file with added parameters (Fig. 2). While a JSON file would work for some applications, we chose CSV because it was supported by *Tinkerplots*. Additionally, CSV is easily manipulable in other more common software tools, such as Microsoft Excel.

### 3.1.3 A Classroom Hub for Wearable Devices

We noted earlier that students may lack access to mobile devices and computers necessary to fully obtain data from a wearable QS device. Furthermore, it is not yet common for there to be 1-1 computing devices in elementary classrooms. In response to this, we prepared a device cart from plastic tackle boxes, USB hubs, and a wheeled shelving unit (Fig. 3). This shelving unit established a clear home for students to charge their devices and enabled clear accounting of devices that went missing (a common occurrence with elementary students). In the context of a classroom with a large number of users, limited time to manage data synchronization and device charging, and a need to move and store equipment quickly and compactly, custom technology carts like this are a simple, relevant, and in some ways, elegant solution that have proven effective for other new classroom learning technologies [35].

We have been able to provide a separate computer that linked to the cart, although in some years, we have used classroom-based machines. When connected to the Internet,



Fig. 3. Custom cart for storing, charging, and syncing a classroom set of 30 Fitbits devices when they are not used by students. Having such a cart is important in classroom contexts where class sets of equipment need to be mobile so they can be transported quickly.

this cart became a wireless communication hub with a USB antenna that could communicate with any Fitbit device in proximity. This one antenna location could eventually communicate with all the Fitbit devices. Because the typical use case predicts no more than a few devices attempting to sync from a given terminal, having 30 devices attempting to sync through a single classroom computer requires the devices be within antenna proximity for longer periods of time. Since students are already in their own classroom for several hours a day, having just this one (or even better is to have a few classroom desktop computers involved) with a hidden antenna running throughout the day in the background can, in principle, support continuous data transfer without requiring students to have their own separate syncing devices. As a failsafe, manual syncing can be forced through the hub cart as well. Altogether, even though it is fairly low-tech, we view the cart as actually an important and critical tool for realistic use of wearable devices and their data in the classroom. Teachers who are charged with monitoring dozens of small devices amidst all their other responsibilities are far less likely to be willing to use such devices as a learning tool. Thus, small equipment hacks like this can determine whether or not a given technology, will be feasible in the classroom.

## 3.2 Classroom Activities

To motivate QS device use, we designed a unit using a two pronged approach to provide students with frequent opportunities to encounter and talk about data. Each prong was included as part of the daily activities, and these were designed for roughly four weeks of classtime. The first prong was an opportunity for students to view and discuss data about their daily activities collected using the Fitbit devices. These were open-ended conversations about the data that students had obtained and involved activities where different visualizations of students' activity data were created or projected in class. These activity data review times were opportunities for students to notice tendencies in their activity data, such as when students in the class tended to be more active. They were also times to raise questions for later student-led investigation.

In parallel, a second prong of activity that took place each day had students collecting more traditional measurement data, such as lengths of common classroom objects (e.g., pencils, lockers, etc.) using canonical and non-canonical measuring devices. The reason for this was to generate some of the same distributional and statistical phenomena across types of measurements (e.g., number of steps and length of an object as two different types of measures), such as variation in distributions, or to highlight common issues involved with data such as the need to obtain many measurements. These activities progressed over three weeks with measurements being collected each week, represented graphically, and compared against other sets of measurements. While inspecting measures together as a class, possible measures of center (e.g., modes, medians, means) were introduced. By the fourth week of the unit, the students were working in small groups to gather two sets of unexamined measurements from two schoolday activities that they found interesting, using the conceptual and physical tools that were introduced in the weeks prior.

The guiding principles motivating these two activity prongs included a desire to privilege familiarity with activities that produced activity data (i.e., using students' recollections of recess activities to understand representations of data from recess) and to move students fluidly between different forms of measurement and representation. Thus, discoveries from daily activity data were used to make sense of more traditional data (i.e., length measurements) and vice versa.

To illustrate how these activities in class, we present next a series of brief vignettes summarized from video we recorded each day from one of the two participating fifth-grade classrooms. Following those, a set of quantitative and comparative learning results are presented.

### 3.2.1 Thinking About Steps Taken During PE

Very early in the unit, a class of fifth graders that was being taught by Mr. Rush (a pseudonym, as are all proper names of people in this article) and assisted by a student teacher, Ms. Buttars, had begun talking about how many steps were being recorded by the Fitbit devices they were each already wearing. This particular vignette, which featured Ms. Buttars asking questions and Mr. Rush taking notes on the whiteboard, was selected because it shows students accessing some of their own knowledge about schoolday activities, namely what happens in physical education class, to begin thinking about the kinds of data they would be collecting.

Ms. B:     Okay, so what did you guys notice [on your Fitbit devices] after you went to PE?
Asher:     We had more.
Liam:      I had 7,000 steps.
Ms. B:     Do you think you'll have more steps yesterday or today?
Students:  Today!
Ms. B:     Why?
Devin:     Because we are running the mile!
Jonas:     And because we started later [yesterday] than we do today.
Ms. B:     So these will be all your steps you take during school. So you guys are running the mile today. What did you guys do in PE yesterday? Did you run pacers?
Sylvia:    Yes, we ran pacers. And we still need to run the mile.
Ms. B:     Do you think you are going to run more steps in the mile than in the pacers?
Students:  (several speaking simultaneously) No! Yes! It depends!

This brief exchange that involved talking about step counts prior to viewing a data visualization represents a "first-prong" activity. This kind of activity built on what students alrady knew about daily activity and used Fitbit data. The students knew what activities took place during PE and on which days (the school had a separate PE teacher for all grades) and what they knew about PE was being privileged in order to help think about their data.

This excerpt includes extra details that the students remembered, such as the observation that they started PE late on the previous day (see Jonas's comment). While it may seem extraneous or a routine observation, it was still
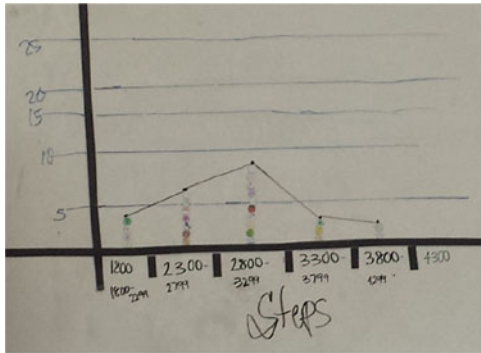
Fig. 4. Data plot generated by the class using magnets on the whiteboard showing how many steps were recorded for students during PE in which they had to do a mile run.

notable that a student recognized how a different start time could actually affect the number of steps that could be taken or recorded. First-prong discussions were intended to privilege such observations so that students would be equipped to think critically about their data by reflecting on how data had been collected.

Additionally, Ms. B asked the students if pacers or the mile would get more steps. The students were uncertain. We learned that this was because both pacers and running the mile were known as demanding PE activities. This was noteworthy because it showed that intuitions about what was more demanding were starting to be accessed and evaluated. It also shows compactly how quantifiable questions about the school day could be raised organically.

In all, this brief snippet serves to demonstrate that even briefly reflecting on data that they had obtained enabled students to begin building on what they already knew and to pose questions that they could potentially answer with QS technology.

### 3.2.2   Discussing Sources of Variability in Steps

Despite the opportunity, the students did not go on to examine step data on pacers. Teachers must make choices about what comments to build upon over the course of a curricular unit, and pacers were not chosen. However, the next time the class convened, the students had gotten their number of steps during that next PE day (after they had run the mile). Per the design for the day's activity, the class worked together to produce a histogram of their data (Fig. 4).

For this histogram, each student had a single round magnet that they used to place their own number of PE steps onto a magnetic whiteboard at the front of the class. Once the data were plotted, a roughly normal distribution appeared. The teachers then asked the class to talk about what they noticed and what were possible sources of variability.

Ms. B:      What do you notice about your steps?
Sylvia:     Most people had between 2,800 and 3,299.
Mr. R:      Tell me more about your information. What could make the difference? Why isn't there everybody taking the exact same number of steps? I need to know the answer to that, Brendan.
Brendan:    Small or large.
Mr. R:      Small or large what?

Brendan:    People.
Kyle:       Legs.
Mr. R:      Small and large legs. Small and large people. Okay, so people's height may make a difference. What else?
Carrie:     Their steps. They might be smaller.
Mr. R:      Okay, so maybe a person's steps might be different.

This proceeded to discussions about how students could also have gone a little farther in their mile runs since the number of steps recorded may not have been cut off exactly at the end of their miles. However, rather than cover all aspects of that later discussion, we wish to use this second short vignette to highlight a few things. One is a strategic move we had made to make the students' later examination of data in *TinkerPlots* more intuitive by having each data point represented by a dot similar to what was shown in the software. As students began to encounter data later, they would stick the circle magnets corresponding to their data on a projected *TinkerPlots* display that had data hidden. Since this, and other classes where we have worked, have magnetic whiteboards that can and often do double as projection screens, this was feasible. With magnets already in place on the whiteboard, the data on *TinkerPlots* would be revealed so that the class could see how their manual plotting compared to what *TinkerPlots* automated and gradually move on to using *TinkerPlots* exclusively to produce data visualizations.

This vignette also highlights how measures of center were planned to also naturally emerge through student talk about their QS data. For instance, the mode is a measure of where there are the most values in a distribution. Rather than having students learn strictly the computational procedure up front (i.e., count all instances of different values, rank the number of instances, and report the value that has the most instances), we planned for the class to situate what seemed intuitively representative or typical in the data displays they created. In subsequent days, these were formalized with official names and procedures. In this particular vignette, Sylvia made an immediate observation of where the "most people" were in terms of their steps, which was setting up recognition of the mode and how it appeared in their display of data.

Finally, this excerpt has a brief record of student and teacher talk about sources of variability. As this was identified as an important topic to address in previous design research, we note here that by using steps, students had intuitions about what led to variability. In this case, Brendan, Kyle, and Carrie could talk about some possible sources of variability that they were already knew well (e.g., height, step size, etc.). This idea of variation could then be extended to the other kinds of measurements that they had obtained, which appears in the next vignette.

### 3.2.3   Seeing Variation Across Different Data Sets

Continuing with the data obtained from PE and the mile run, this next vignette comes from a later part of the unit and serves to show commonalities in data distributions. We had mentioned earlier that the "second-prong" of our designed activities involved using other measures besides steps, such as length, in conjunction with Fitbit device-computed steps
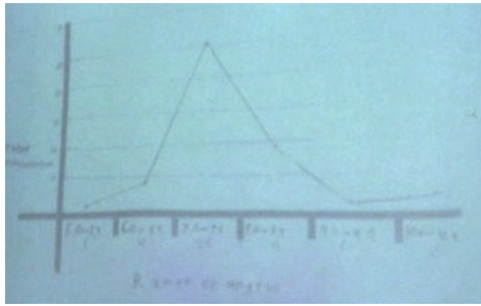
Fig. 5. Line shape of distribution from a previously produced histogram showing pencil measurements that was photographed and projected on the whiteboard alongside the distribution in Fig. 4. Obtained from video screen capture.
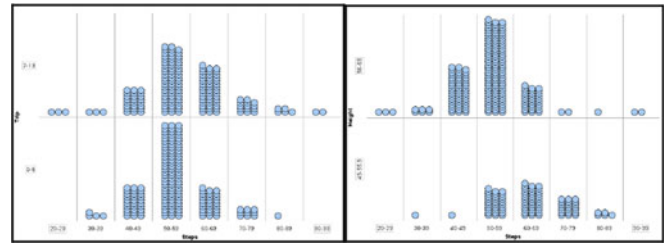


Fig. 6. Figure showing TinkerPlots rendered data plot that had multiple distributions of student steps across a playfield. The comparison on the left shows steps taken later (top) and earlier (bottom) in the day. The comparison on the right used the same data but was divided by height with taller students on top.

as a form of measurement. This was to help students understand variation as a natural and ubiquitous phenomenon in a data distribution. The excerpt below comes from when the class is shown a projected image of pencil measurements that they had obtained and the same plotted PE data (Fig. 5). Mr. R and Ms. B asked the students if they saw things in the two plots that were similar or different.

Zack: You see the one on the board, you count from the left and to the 2,800 – 1, 2, 3 [bins] and the same thing with that (points to projected image) if you count from the left until it goes to the highest. 1, 2, 3 [bins].

Ms. B: So you're saying the peak comes up to the same spot in the middle. They're not the same numbers, because the numbers are a lot smaller.

Zack: They're not the same numbers but the peak comes at the same spot of the highest

Ms. B: So the peak comes up, not quite in the middle but close. Abe?

Abe: Well, I see something that's similar in both of them that's like, when it keeps getting toward the middle, it keeps getting higher. So maybe the people that got somewhere in the middle had the most steps – um, measurements.

Ms. B: So we had more people that got those?

Abe: Yes. Toward the middle, most of the people had in both charts, had the same.

Through this excerpt, two students in the class discussion observed similarities in two of the first distributions. Zack noticed a similar peak that was located in roughly the same positions in both plots. As was the case for many students, he found it easy to notice the mode. Although it was a brief encounter, Zack was also beginning to recognize, with prompting from Ms. Buttars, that such a tendency could be identified within distributions regardless of the magnitude of the quantities that were being measured ("They're not the same numbers but the peak comes at the same spot"). Abe also recognized this point but also noted that beyond having the highest peak, the stacks of data points would be higher the closer they were to the bin with the highest peak. This was the beginning of recognizing and becoming comfortable with distributional shapes (e.g., normal curves). When coupled with the earlier observations about sources of variation, this could be seen as laying a

foundation for better understanding of how data could be graphically represented, what measures of center actually reference in a set of data, and how variability is to be expected and depicted.

### 3.2.4 Comparing Two Distributions as a Class

After obtaining several different measurements as a class that included step data and size measurements and pooling their data, the class proceeded to look at two different distributions. One example of this was trying to figure out if the steps taken by students at later times in the day differed from those taken at earlier times across a common playing field outside their classroom (a comparison that the students themselves had suggested). They then used the same dataset working in small groups to compare the number of steps taken by tall students and short students (defined as taller or shorter than the median height) (Fig. 6). Given the limited space, we do not provide transcript excerpts but rather report that we observed that other groups opted to use measures of center and qualitative descriptions of the distributions to describe differences too, comparable to what was illustrated in excerpts before. This suggests they were indeed building off their intuitive ways of talking about data, as we had hoped and intended.

### 3.2.5 Small Group Quantified Selves Projects

Following this opportunity to look at comparisons of data, the students then broke into small groups and then proposed their own Quantified Selves projects. For those projects, the students were to use step data obtained from the Fitbit devices every student in the class was wearing. If the data did not already exist for a specific activity, they could propose that a group of students or the entire class do some activities to generate necessary data.

To illustrate what these projects looked like, we share one group's QS project. This particular group was interested in comparing how active the game "Capture the Flag" was relative to "Ball Tag," which involved the person playing "It" tagging another player by hitting them with a playground ball (Fig. 7). For the students, being more active was operationalized by having more steps per minute.

What they found was that while the spread of the data was similar between the two games, there was a more normal distributional shape when playing ball tag. It also had a slightly higher average and median, although capture the flag had a higher mode. The reason for this, they inferred,
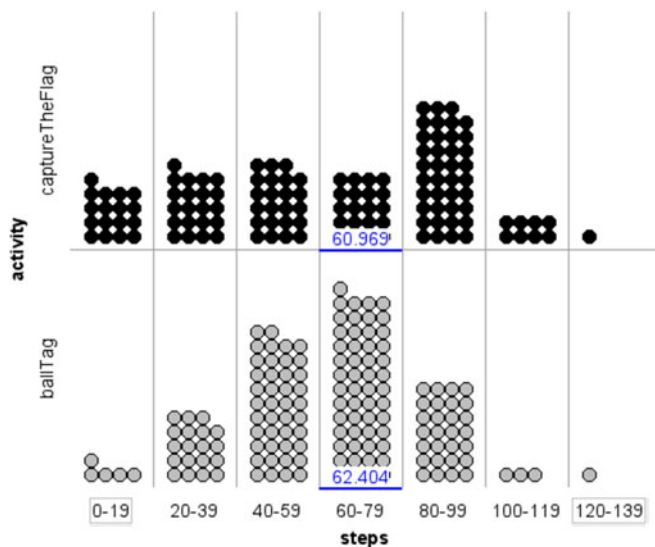
Fig. 7. Frequency plot of data collected from 18 students playing capture the flag and ball tag for 20 minutes per game. Each dot represents one minute of gameplay for one student. The numbers superimposed on the plots represent the mean.

was that capture the flag has a period of time when there is intense running by players on both teams to steal or recover the flag. Other students pursued similar projects and made comparable observations.

## 4 ASSESSMENT INSTRUMENT AND PROCEDURES

Having briefly summarized how QS technologies were integrated into the classroom, we turn now to assessment. To assess student thinking after the designed unit was completed, we used a new, psychometrically validated statistical reasoning assessment instrument developed by researchers at Vanderbilt University and the University of California, Berkeley [36]. This new assessment instrument was founded on the idea of Learning Progressions, a construct popularized by a consensus report published by the National Research Council [37] and also in special issues of major education research journals [38]. The basic idea of a learning progression is that it represents a possible trajectory for student learning of sophisticated disciplinary content and practices in a manner that revisits itself over time. Rather than specify some content that should be understood at some grade level (as if knowing is a binary state), learning progressions focus on *levels* of understanding of particular content and student facility with specific disciplinary practices. It is also informed by current research in cognitive science and design-based educational research that has demonstrated how far students can progress in areas such as science and mathematics when given adequate tools and a supportive learning environment. Learning progressions can not only serve assessment aims but also be a guide for designing new instructional materials and tools [36]. Midway through the enactment of the unit in Mr. Rush's class, we learned of the availability new of the instrument and were able to obtain a copy.

The instrument we used focused specifically on a new learning progression on statistical reasoning and had nine multi-part items related to five constructs, summarized below:

1. Conceptions of Statistics (CoS) – Ability to describe and characterize distributions and measures of center.
2. Data Display (DaD) – Ability to create and read canonical data representations such as density plots and histograms.
3. Metarepresentational Competence (MRC) – Ability to critically evaluate alternate forms of data representation relative to what is to be communicated and who is to be reading a data representation.
4. Informal Inference (InI) – Ability to draw inferences from multiple sets of distributional data based on overall appearance and assuming inherent variability.
5. Modeling Variability (MoV) – Ability to understand how changes in measurement systems will produce variability and the inherent variability associated with more and less precise measurement approaches.

Each construct had between three and eight levels of performance with the lowest levels being akin to having little to no formal instruction (e.g, when asked to visually show information from a word problem, students may literally draw a picture of the scenario and list numbers in the margin, a response we did receive from some students) and highest levels demonstrating a sophisticated understanding akin to what we might associate with a data savvy high school student (e.g., students create an organized histogram or density plot of data with standardized bin sizes and include details about center and spread). The items were often quite complex in their demands, but we relied on established and validated rubrics to evaluate student answers.

To illustrate the nature of the assessment, consider an item like the following:

Emma is a fan of the Telluride Prospectors, a minor league baseball team in her hometown. Emma tracked the number of runs they scored in their 30 home games. The list below shows how many runs they scored in each of their games.

1, 2, 3, 6, 3, 4, 3, 6, 8, 4, 8, 4, 5, 5, 2,
6, 5, 5, 4, 3, 3, 11, 4, 2, 0, 9, 4, 4, 1, 0

Given these scores, make a display that helps you think about how many runs you expect the Prospectors to usually score during a home game.

An item like this assesses students' competences with Data Displays, as the student would ideally create a display that organizes the numbers in a way such that they can identify a center or concentration of values, and use that to provide a written prediction and justification for that prediction. If this item had a comparison data set (such as how many runs are scored in away games), and the student were asked to determine if there were a difference in their performance and to articulate their basis for making that judgment, it would be appropriate as an Informal Inference task as well.

Students had one class period during which to complete the assessment. They completed most of the assessment on a computer and also submitted a paper portion where they included any data displays that they had created for the assessment. When we were able to perform a pre-post test comparison, the items on the pre and post were matched so that the students were being assessed with the same tasks.
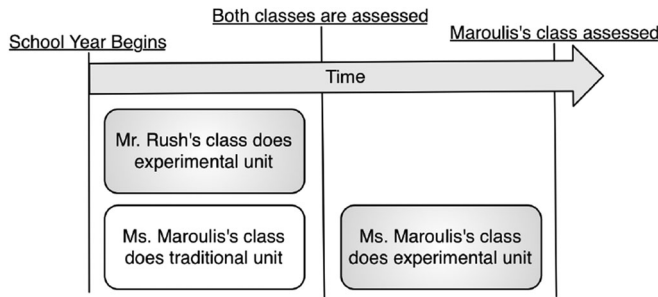
Fig. 8. The unit deployment schedule showing the relative unit start and assessment times for each class.

A single rater scored the items after he had obtained over 80 percent reliability on rating open-ended items on a subset of responses when compared with an independent rater.

## 5  COMPARATIVE STUDY DESIGN

One class, led by Mr. Rush, began our experimental unit first at the beginning of the school year. Excerpts from his class were provided in the vignettes above. During this same period of time, the other fifth-grade classroom at the school, led by Ms. Maroulis, proceeded with her regularly scheduled instruction. Ms. Maroulis had planned for students to do the standard instruction related to data and measurement based on existing textbook materials and consistent with what she had successfully taught before. This involved covering procedures for computing measures of center and making class pictographs, bar graphs, and pie charts to show data.

As Mr. Rush's class was finishing their unit, Ms. Maroulis then began the designed unit with her class. At that point in time, both classes were assessed using the written test instrument. This simultaneous assessment allowed us to produce a small quasi-experiment comparing how standard instruction compared against an experimental approach using the previously described technologies and activities. Following their own completion of the wearable device unit, the students in Ms. Maroulis's class were assessed again using the same instrument to see how their post-unit performance compared against that of the students in Mr. Rush's class (Fig. 8). Because we did not have access to the test prior to Mr. Rush's enactment, we do not have a record of his class's prior performance. However, the students were assigned to each teacher during the summer by the school's administration with the goal of making sure the populations were balanced. For this reason, we do not suspect there to have been substantial performance differences between the two groups at the beginning of the school year when Mr. Rush's class first began the designed unit. However, we cannot be certain there was no population difference prior to our assessment, and thus the results should be considered as demonstrating potential of our approach rather than stating its relative efficacy compared to others.

## 6  QUANTITATIVE RESULTS

Overall, when comparing Mr. Rush's class performance (who had just completed the newly designed, wearable device emphasizing unit) against that of Ms. Maroulis
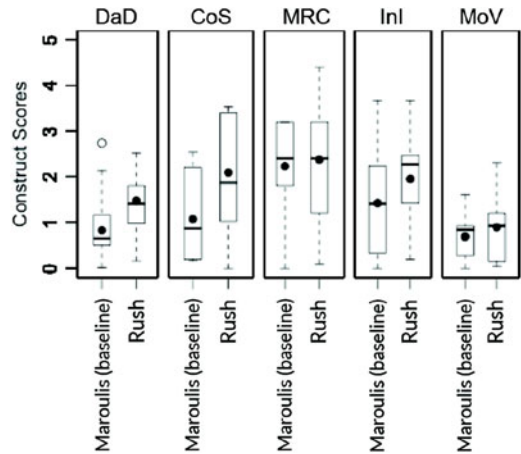


Fig. 9. Ms. Maroulis's students' pre-test scores compared with Mr. Rush's students' post-test scores. In this comparison, Ms. Maroulis's students are being used to benchmark the progress of Mr. Rush's students.

(who taught elementary statistics content traditionally), a significant difference in the performance of students appears across two constructs at the 0.05 level (Fig. 9). These were Conceptions of Statistics ($t = -3.71, df = 47, p < 0.001$) and Data Display ($t = -3.17, df = 42, p = 0.01$). There was no significant difference on performance with MetaRepresentational Competence ($t = -0.65, df = 46, p = 0.52$) nor on Modeling Variability ($t = -1.34, df = 43, p = 0.19$). There is a trend toward greater performance on Informal Inference ($t = -1.96, df = 47, p = 0.06$) (Table 1).

To test these apparent differences further, we looked at the same baseline scores from Ms. Maroulis's class before the unit and at their scores on the same assessment after they had completed the wearable-supported unit (Fig. 10). It turned out a similar pattern appeared in comparing the pre-wearable unit and post-wearable unit test scores for students in Ms. Maroulis' class.

There were significant differences in test scores for three constructs at the 0.05 level in Ms. Maroulis's class. These were Conceptions of Statistics ($t = -2.57, df = 23, p < 0.05$), Data Display ($-7.21, df = 23, p < 0.001$), and Modeling Variability ($t = -2.11, df = 23, p < 0.05$). Again, there was no significant difference on performance with MetaRepresentational Competence ($t = -0.61, df = 23, p = 0.55$). Also, as with Mr. Rush's class, there was a trend toward gains in Informal Inference ($t = -1.84. df = 23, p = 0.08$) (Table 2).

## 7  DISCUSSION

In looking at these results, we see similar (although not identical patterns) in student performance. From using the QS technologies and the instructional sequence we

TABLE 1
Comparison of Maroulis Pretest and Rush Post Test

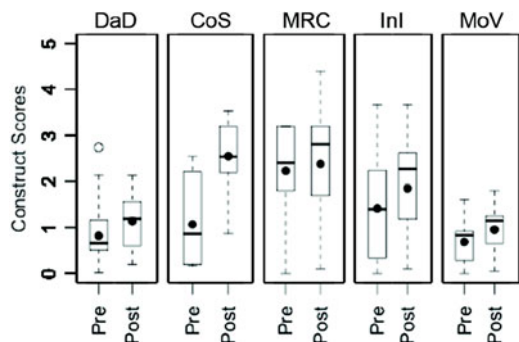| Construct | Maroulis Preset | | Rush Post test | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| CoS | 0.822 | 0.652 | 1.469 | 0.621 |
| DaD | 1.068 | 0.885 | 2.088 | 1.229 |
| MRC | 2.225 | 1.107 | 2.371 | 1.321 |
| InI | 1.418 | 1.046 | 1.949 | 0.907 |
| MoV | 0.688 | 0.491 | 0.888 | 0.660 |

Fig. 10. A comparison of Ms. Maroulis's students' pre- and post-test scores, showing significant differences in DaD, CoS, and MRC.

TABLE 2
Comparison of Maroulis Pretest and Post Test

| Construct | Maroulis Preset | | Rush Post test | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| CoS | 0.822 | 0.652 | 1.1425 | 0.610 |
| DaD | 1.068 | 0.885 | 2.5417 | 0.729 |
| MRC | 2.225 | 1.107 | 2.375 | 1.335 |
| InI | 1.418 | 1.046 | 1.850 | 1.056 |
| MoV | 0.688 | 0.491 | 0.954 | 0.507 |

designed, we saw significant differences in Conceptions of Statistics. This meant that learners' understanding of the core terminology of elementary statistics, and of the procedures by which certain standard values (e.g., ranges, medians) are computed, improved. Interestingly, even though Ms. Maroulis covered these topics in the traditional way while Mr. Rush's class used the new technologies, her students still showed improvement in this area when they had a chance to use the technologies themselves. This is not to say that QS activity tracking inherently supports improved learning of statistics, as a range of factors could be at play such as the designed activities and the teachers' facilitation. However, it seems plausible especially given the qualitative examples above that noticing and talking about measures of center with familiar activity data served as leverage. We have seen evidence of this elsewhere [39]. At a minimum, we appear to have a case of students making greater gains in understanding elementary statistics conceptions than with traditional instruction.

The largest gains are in Data Display. As density plots and histograms are especially common in statistics but are only minimally touched upon in elementary classroom instruction, these gains are encouraging. Again, it is hard to pinpoint exactly what led to this improvement as there are a myriad of possible factors. However, it is our hope that our approach of having students convert records of their own physical activity into physical data points, and also helping students to see how those data could be arranged and organized physically (with magnets) and virtually (in *TinkerPlots*), helped in this regards. Our observations of students working with both magnet and automatically rendered plots suggest that may have been so.

MetaRepresentational Competence had no significant difference for either group. This could be for a number of reasons, not least of which the students were just not given sufficient exposure to these ideas. As shown in our observational excerpts, the unit we prepared did not explicitly stress much reflection as to why particular forms of data representation are more or less useful for showing features of physical activity. In addition, this was an area in which students in Ms. Maroulis's class performed the highest the first time they were assessed. Their post-unit scores were similar, based on visual appraisal, suggesting it could also be a ceiling effect. MetaRepresentational Competence is still a relatively new construct in educational research [40] and ways of supporting its growth are still being identified.

Regardless, this is an area where we see still more room for improvement.

Informal Inference performance was more surprising given the experiences with comparison of distributions that we have described earlier. With a firm alpha level cut-off of 0.05, the students did not show significant differences from the control classroom (i.e., Ms. Maroulis's class just after they completed their traditional approach to learning elementary statistics). The opportunities that students had to do comparisons and talk about differences between two data sets generally came about in whole class discussions and during the quantified selves-like projects that they pursued. It could be that those projects, while catering toward personal student interests, were not adequately scaffolded for students to become more articulate about how they distinguished between distributions.

Finally, Modeling Variability is the one construct where one class showed a significant difference and the other class only suggested a trend. Again, the values are fairly similar, although the mean is higher for students in Ms. Maroulis's class. This could have been due to slight differences in how each teacher ran their lessons, or it could be that the black-box nature of wearable activity trackers did not allow students to give more consideration to how those devices, as measurement instruments, are still prone to reporting values that have some error. We have seen instances of this being a difficulty for students unless it is explicitly addressed and explored in the classroom [41], and it may be that future iterations of QS activity tracker data analysis in school settings needs to better incorporate discussions of how to formally describe error and variability in the data produced by the devices.

## 8 CONCLUSION

Through this article, we have taken seriously the possibility that some of the newly developed wearable technologies associated with the QS movement, and specifically those that track physical activity, could have a productive place with elementary school classrooms. As with any new learning technology that is to be used by a classroom and with younger students, a number of considerations had to be made and new solutions had to be devised to make the devices and data usable and accessible. However, we did succeed in finding solutions (that were also greatly informed by the existing research literature related to learning technologies and statistics education).

Like in previous work, we have gone beyond the standard health and goal-oriented use cases of wearable activity tracking technology. Yet in contrast to earlier papers, this

article is showing performance differences across whole classes of students and the kinds of classroom conversations that supported those changes. It is also our first reporting of a classroom-level intervention given the new modifications and tools we have created, such as the data grabber and hub cart.

Currently, we are at a point with advances in wearable technology are becoming more pervasive and affordable. Still, we recognize they may not yet be affordable for all classrooms or all students. The example intervention shown here is intended to describe one possible future for wearable devices in education and also to show how having students self-quantify and draw on their own knowledge of what routine activities can be represented with data could be beneficial for some statistical content learning. Self-quantification could still involve other mobile devices or even manual recording. When one looks broadly at self-tracking as a practice, the kinds of technologies that people who self-quantify use actually show a great deal of variability [3]. Wearable commercial devices are perhaps the most visible of those right now. They have helped establish a popular foothold for use of data from one's routine experience in new ways. As more design and development work around wearable and self-tracking technologies and practices is pursued in the future, we hope that our field will be in a place to disentangle the extent to which the Quantified Self movement will have a beneficial place in the classroom.

## ACKNOWLEDGMENTS

## REFERENCES

[1] V. R. Lee, "The quantified self (QS) movement and some emerging opportunities for the educational technology field," *Educ. Technol.*, vol. 53, no. 6, Nov./Dec. 2013, Art. no. 39.

[2] V. R. Lee and J. Drake, "Quantified recess: Design of an activity for elementary students involving analysis of their own movement data," in *Proc. 12th Int. Conf. Interaction Des. Children*, 2013, pp. 273–276.

[3] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, "Understanding quantified-selfers' practices in collecting and exploring personal data," in *Proc. 32nd Annu. ACM Conf. Human Factors Comput. Syst.*, 2014, pp. 1143–1152.

[4] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 557–566.

[5] C. C. Ching and S. Schaefer, "Identities in motion, identities at rest: Engaging bodies and minds in fitness gaming research and design," in *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*, Abingdon-on-Thames, U.K.: Routledge, 2014, pp. 201–219.

[6] *Next Generation Science Standards*, Appendix F. Achieve, Inc., Washington, D.C., 2013.

[7] *Common Core State Standards*, CCSS.Math.Content.6.SP.A.1-B.5, National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington, D.C., 2010.

[8] L. Lyons, "Exhibiting data: Using body-as-interface designs to engage visitors with data visualizations," in *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*. Abingdon-on-Thames, U.K.: Routledge, 2014, pp. 185–200.

[9] V. Lee and J. Thomas, "Integrating physical activity data technologies into elementary school classrooms," *Educ. Technol. Res. Dev.*, vol. 59, no. 6, pp. 865–884, Dec. 2011.

[10] V. R. Lee, J. Drake, J. Thayne, and R. Cain, "Opportunistic uses of the traditional school day through student examination of Fitbit activity tracker data," in *Proc. Interaction Des. Children*, 2015, pp. 209–218.

[11] T. White, A. Booker, C. C. Ching, and L. Martin, "Integrating digital and mathematical practices across contexts: A manifesto for mobile learning," *Int. J. Learn. Media*, vol. 3, no. 3, pp. 7–13, Jul. 2011.

[12] H. Beyer and K. Holtzblatt, *Contextual Design Defining Customer-Centered Systems*. San Francisco, CA, USA: Morgan Kaufmann, 1998.

[13] K. Bødker, F. Kensing, and J. Simonsen, *Participatory it Design Designing for Business and Workplace Realities*. Cambridge, MA, USA: MIT Press, 2004.

[14] J. P. Smith, A. A. diSessa, and J. Roschelle, "Misconceptions reconceived: A constructivist analysis of knowledge in transition," *J. Learn. Sci.*, vol. 3, no. 2, pp. 115–163, 1994.

[15] V. R. Lee and M. Briggs, "Lessons learned from an initial effort to bring a quantified self 'Meetup' experience to a new demographic," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Adjunct Publication*, 2014, pp. 707–710.

[16] J. Cai, J. J. Lo, and T. Watanabe, "Intended treatments of arithmetic average in U.S. and Asian school mathematics textbooks," *School Sci. Math.*, vol. 102, no. 8, pp. 391–404, Dec. 2002.

[17] W.-M. Roth, M. K. McGinn, and G. M. Bowen, "How prepared are preservice teachers to teach scientific inquiry? Levels of performance in scientific representation practices," *J. Sci. Teach. Educ.*, vol. 9, no. 1, pp. 25–48, Feb. 1998.

[18] J. M. Watson and J. B. Moritz, "The development of concepts of average," *Focus Learn. Problems Math.*, vol. 21, no. 4, pp. 15–39, 1999.

[19] J. Mokros and S. J. Russell, "Children's concepts of average and representativeness," *J. Res. Math. Educ.*, vol. 26, no. 1, pp. 20–39, Jan. 1995.

[20] V. Dubreil-Frémont, C. Chevallier-Gaté, and N. Zendrera, "Students' conceptions of average and standard deviation," in *Proc. 9th Int. Conf. Teaching Statist.*, 2014.

[21] A. Pollatsek, S. Lima, and A. D. Well, "Concept or computation: Students' understanding of the mean," *Educ. Studies Math.*, vol. 12, no. 2, pp. 191–204, May 1981.

[22] S. Strauss and E. Bichler, "The development of children's concepts of the arithmetic average," *J. Res. Math. Educ.*, vol. 19, no. 1, pp. 64–80, Jan. 1988.

[23] A. J. Petrosino, R. Lehrer, and L. Schauble, "Structuring error and experimental variation as distribution in the fourth grade," *Math. Thinking Learn.*, vol. 5, pp. 131–56, 2003.

[24] C. Konold and A. Pollatsek, "Data analysis as the search for signals in noisy processes," *J. Res. Math. Educ.*, vol. 33, no. 4, pp. 259–289, Jul. 2002.

[25] R. Lehrer, "Learning to reason about variability and chance by inventing measures and models," presented at the *Moore Lecture Series, Friday Institute for Educational Innovation, College of Education, NC State University*, Raleigh, NC, Nov. 2007.

[26] B. Hug and K. L. McNeill, "Use of first-hand and second-hand data in science: Does data type influence classroom conversations?," *Int. J. Sci. Educ.*, vol. 30, no. 13, pp. 1725–1751, Oct. 2008.

[27] U. Wilensky and W. M. Stroup, "Networked gridlock: Students enacting complex dynamic phenomena with the HubNet architecture," in *Proc. 4th Annu. Int. Conf. Learn. Sci.*, 2000, pp. 282–289.

[28] D. Abrahamson and U. Wilensky. "S.A.M.P.L.E.R.: Collaborative interactive computer-based statistics learning environment," in *Proc. 10th Int. Congr. Math.* 2004.

[29] P. Cobb and C. Tzou, "Learning about data analysis," in *Mathematical Representation at the Interface of Body and Culture*, Charlotte, NC, USA: IAP, 2009, pp. 135–170.

[30] C. Hancock, J. J. Kaput, and L. T. Goldsmith, "Authentic inquiry with data: Critical barriers to classroom implementation," *Educ. Psychol.*, vol. 27, no. 3, pp. 337–364, Jun. 1992.

[31] C. Konold and C. D. Miller, *TinkerPlots*. Emeryville, CA, USA: Key Curriculum Press, 2005.

[32] C. Konold, "Designing a data analysis tool for learners," in *Thinking with Data*, M. Lovett and P. Shah, Eds. New York, NY, USA: Lawrence Erlbaum Associates, 2007, pp. 267–292.

[33] W. Finzer, *Fathom Dynamic Data Software*. Emeryville, CA, USA: Key Curriculum Press, 2005.

[34] J. Takacs, C. L. Pollock, J. R. Guenther, M. Bahar, C. Napier, and M. A. Hunt, "Validation of the Fitbit one activity monitor device during treadmill walking," *J. Sci. Med. Sport Sports Med. Australia*, vol. 17, no. 5, pp. 496–500, Sep. 2014.

[35] C. McKay and K. A. Peppler, "MakerCart: A mobile fab lab for the classroom," presented at the *Interaction Design for Children*, New York, NY, USA, 2013.

[36] R. Lehrer, M.-J. Kim, E. Ayers, and M. Wilson, "Toward establishing a learning progression to support the development of statistical reasoning," in *Learning over Time: Learning Trajectories in Mathematics Education*, Charlotte, NC, USA: Information Age Publishers, 2014.

[37] R. A. Duschl, H. A. Schweingruber, and A. W. Shouse, Eds., *Taking Science to School: Learning and Teaching Science in Grades K-8*, 1 ed.. Washington, D.C., USA: National Academies Press, 2007.

[38] R. G. Duncan and C. E. Hmelo-Silver, "Learning progressions: Aligning curriculum, instruction, and assessment," *J. Res. Sci. Teach.*, vol. 46, no. 6, pp. 606–609, 2009.

[39] V. R. Lee and M. DuMont, "An exploration into how physical activity data-recording devices could be used in computer-supported data investigations," *Int. J. Comput. Math. Learn.*, vol. 15, no. 3, pp. 167–189, Dec. 2010.

[40] A. A. diSessa, "Metarepresentation: Native competence and targets for instruction," *Cogn. Instr.*, vol. 22, no. 3, pp. 293–331, Sep. 2004.

[41] V. R. Lee, J. Drake, and K. Williamson, "Let's get physical: K-12 students using wearable devices to obtain and learn about data from physical activities," *TechTrends*, vol. 59, no. 4, pp. 46–53, Jun. 2015.

**Victor R. Lee** is an associate professor of Instructional Technology and Learning Sciences at Utah State University. His academic research focuses on new uses for wearable technologies to support learning and research, the Quantified Self movement, and the Maker Movements as they apply to formal and informal educational settings.

**Joel R. Drake** is working toward the doctoral degree in Instructional Technology and Learning Sciences Department, Utah State University. His current research interests include informal STEM learning through participation in interest-driven communities of practice.

**Jeffrey L. Thayne** received the doctoral degree in Instructional Technology and Learning Sciences, Utah State Unversity. His current research interests include student motivation and engagement, and technology and design solutions for helping learners connect more to learning activities and content.