Optimizing STEM Education With Advanced ICTs and Simulations

Ilya Levin
Tel Aviv University, Israel

Dina Tsybulsky
Tel Aviv University, Israel
From Wearing to Wondering: Treating Wearable Activity Trackers as Objects of Inquiry

Joel R. Drake  
Utah State University, USA

Ryan Cain  
Utah State University, USA

Victor R. Lee  
Utah State University, USA

ABSTRACT

Wearable technologies represent a rapidly expanding category of consumer information and communications technologies. From smartwatches to activity tracking devices, wearables are finding their way into many aspects of our lives, changing the way we think about ourselves and the world around us. The rapid adoption of these tools in everyday life hints at the possibilities these devices may hold in school and other educational settings. Drawing on examples taken from a five-year study using wearable fitness tracking devices in elementary and middle school classrooms, this paper presents two examples of how wearable devices can be appropriated for use in school settings. These examples focus on instances where students turned activity trackers into objects of inquiry using data from familiar activities.

DOI: 10.4018/978-1-5225-2528-8.ch001
INTRODUCTION

Wearables represent a rapidly growing category of information and communications technologies (CCS Insight, 2015). From smartwatches to activity tracking devices, wearables unobtrusively capture and collect large amounts of data relating to aspects of wearers’ experiences that were previously unavailable. Using sensors, like accelerometers, wearables can quantify a user’s activity (e.g., steps, sleep, breathing) and make it available for inspection. The subsequent analysis of these data can change the user’s sense of self and their relation to the world around them (Lee & Drake, 2013a). Prior to the introduction of wearable technologies, these data required active intervention on the part of the individual to capture and track relevant data—manually measuring distances traveled, logging places visited, etc.

Given the potential wearables have for producing personally-relevant data and their increasing social recognition, it is only a matter of time before these devices find their way into school and classroom settings. To use these devices to their full potential, teachers and researchers must work to understand the opportunities and challenges presented by using these devices in schools. For multiple years, the authors (along with a team of researchers, teachers, and designers) have worked with 5th and 6th grade classrooms in the United States to understand how the use of personally-relevant data from wearable activity trackers affects students’ engagement with and appropriation of statistical content and practices. Over the course of the study, the authors developed, tested, and refined a statistics curriculum and video recorded and analyzed classroom interactions in order to examine how students leveraged their familiarity with activity in making sense of data from wearables. We have seen familiar activities inspire students to pursue lines of inquiry, develop inclusion criteria, and provide context for their interpretation of data (Lee, Drake, Thayne, & Cain, 2015).

Other publications by the authors have focused on using students’ own activities as objects of inquiry—how using activity tracker data can help students better understand these activities (e.g., Lee & DuMont, 2010). The aim of this chapter is to examine how students’ use of activity tracker data to better understand the tracker itself—how well activity trackers capture and quantify activity—can be a productive strategy for fostering evidence-based discourse in classrooms. Through presentation of the classroom examples below, the authors argue that students’ knowledge of familiar activities can be used to foster skepticism towards wearable devices that leads to productive inquiry in a statistics unit, including the eventual resolution of that skepticism through evidence-based discourse. Each of the examples shows a different way that students may use activity data to critically examine the functionality of the wearables and how the students resolve their questions. In the first example, students test whether the trackers they are using are accurate enough for use in other
inquiry activities. The second example involves students investigating whether and how well their devices can track a particular activity and how that ability affects their interpretation of previously collected data.

BACKGROUND

In an effort to reverse US students’ underperformance in science, technology, engineering, and mathematics (STEM) topics (OECD, 2014) and increase interest in STEM careers (President’s Council of Advisors on Science and Technology, 2010), researchers and educators developed the Next Generation Science Standards in part to encourage students’ engagement in scientific practices rather than simply memorize facts (Achieve, Inc., 2013). Among the practices encouraged by the NGSS are: 1) Asking questions and defining problems, 3) Planning and carrying out investigations, 4) Analyzing and interpreting data, 6) Constructing explanations, and 7) Engaging in argument from evidence (Achieve, Inc., 2013, Appendix F). Each of these practices is important to the accomplishment of productive inquiry. However, students can struggle with inquiry curricula for a variety of reasons, including difficulties asking appropriate questions, inexperience with school-based inquiry learning, and underdeveloped representational and interpretive skills (Elby, 2000; Kanari & Millar, 2004; Krajcik et al., 1998; Kuhn, 2007). Research in statistics education has also noted that students’ ability to engage in meaningful inquiry requires an understanding of variability as a natural part of measurement (Petrosino, Lehrer, & Schauble, 2003). Some researchers even recommend that variability be the central concept in any statistics or data-based inquiry curriculum (Konold & Pollatsek, 2002).

Familiarity with the data collection processes can be a powerful tool for interpreting data representations (Hug & McNeill, 2008; Nemirovsky, Tierney, & Wright, 1998). This idea of collecting data first-hand is central to both the commercial popularity wearables and the increasing interest they are receiving from researchers. When researchers look to examine the application of wearable technologies to educational settings, their gaze seems to focus on devices like Google Glass and the Oculus Rift (e.g., Bower & Sturman, 2015). The augmented and virtual reality aspects of these devices offer tantalizing glimpses of what education and life may be like in the future. However, this focus on the technological “deep end” overlooks a variety of device capabilities and applications that are currently realizable (e.g., Sandall, 2016). Over the past several years, researchers used wearable devices to help students and educators examine various aspects of their own experiences. Wearing custom temperature sensors, students collected and later analyzed changes in temperature across various spaces throughout the day (Resnick, Berg, & Eisenberg, 2000).
Point-of-view cameras (e.g., GoPro) have been used to help educators reflect on their teaching practices (Sherin, Russ, Sherin, & Colestock, 2011). While wearing custom “thinking tags,” students were able to study the spread of disease through interpersonal contact (Klopfer, Yoon, & Rivas, 2004). Researchers have used wearables and students’ movements to understand broader scientific phenomena including animal foraging systems (Moher et al., 2014; Peppler et al., 2010) and the effects of global warming (Lyons, Silva, Moher, Pazmino, & Slattery, 2013). Capturing larger-scale body movements has been used in teaching geometry (Hall, Ma, & Nemirovsky, 2014) and GIS-related topics (Taylor & Hall, 2013).

With their more limited feature set, wearable activity trackers are often overlooked when the future of wearable devices in education. However, their relatively lower price point coupled with their ability to provide personally-relevant, health-related data has led to their rapid adoption by adults desiring to track their health (Choe, Lee, Lee, Pratt, & Kientz, 2014; Lee & Drake, 2013a). Noting the popularity of activity trackers and the obesity epidemic, researchers have begun to look into ways that youth can use wearables to better understand their own health (Ching & Schaefer, 2014; Norooz, Mauriello, Jorgensen, McNally, & Froehlich, 2015).

**Wearable Activity Trackers in This Study**

For this study, the authors used wearables created by Fitbit, which are essentially internet-connected pedometers; they count steps and use that information to estimate other activity-related quantities (e.g., distance traveled, calories burned). Unlike standard pedometers, these activity trackers can provide step counts in minute-level increments. The Fitbit Ultra (Figure 1, left), used in early implementations, was designed to be worn at the waist. An on-device LCD display showed a real-time step count. The Fitbit Flex (Figure 1, center), used in the last year of the study, was worn on the wrist and had a five-LED display to indicate progress toward the 10,000-step goal. Like other activity trackers, these devices passively sync activity data to a web service via wireless connection to a smartphone or personal computer. Users can interact with their data via a “dashboard” at fitbit.com (Figure 1, right) or via mobile app.

The authors focused on steps as the primary unit of measurement for the students. Steps are familiar, countable events and are already tracked by a wide number of wearable and mobile devices. Steps can be easily translated into a measure of length or a rate (e.g., steps per minute). However, steps are impermanent and cannot be reexamined unless tracked.
Modern activity trackers use three-axis accelerometers or gyroscopes to sense movement. Every movement is tracked as “activity.” Proprietary algorithms compare the motion signature detected to the shock profile of a step (Figure 2). Movements that match this profile are counted as steps. The accuracy of the devices is based on how reliably the algorithm can differentiate between steps and non-step movements.

Figure 2. Acceleration profile of three steps measured by a waist-mounted accelerometer
Adapted from Khandelwal & Wickstrom, 2016.
Fitbit devices have been shown to track steps accurately in laboratory tests (Diaz et al., 2015; Takacs et al., 2014).

**Wearable Curriculum Implementation**

The curriculum included frequent opportunities for students to explore data from their routine activities (Lee et al., 2015). These data explorations typically included students viewing data representations, discussing as a class the meaning of the data represented (e.g., typicality, variation, and shape) and comparing features of the representation to other datasets. Because standard dashboard for these trackers is not intended for such analysis, the authors developed a PHP web form that retrieved data from the Fitbit servers in one-minute increments. These data were returned to the user as comma-separated values (CSV). The CSV was then imported into Tinkerplots (Konold & Miller, 2005) software for data visualization and analysis. Tinkerplots was designed specifically for elementary and middle grade students to create and interpret data visualizations. Using a drag-and-drop interface, students can quickly organize and reorganize data in ways that make sense to them. For this study, the authors focused students’ attention on understanding and interpreting time-ordered and frequency distributions.

These data explorations were supplemented with instruction on statistical concepts (e.g., measures of center, variability) to help students concretize their comparisons. Through these discussions and the supplementary instruction, the students came to question their assumptions about the nature of their daily activities. In some instances, students’ investigations led them to explore the functions of the devices themselves. It is these latter instances that the authors wish to explore in this chapter.

*Figure 3. Time-ordered (l) and frequency distribution (r) for one student’s recess data; each dot represents one minute of recess activity.*
From Wearing to Wondering

METHODS

The data for this study comes from three years of a design research study (Brown, 1992) examining how students could leverage their familiarity with their own activities while learning and applying statistical techniques to interpret data collected during those activities. Over these three years, the curriculum was iteratively refined between enactments in six classrooms. These enactments averaged 24, approximately hour-long sessions. Classroom interactions during each session were captured using two video cameras resulting in over 145 hours of video.

To analyze these videos, the research team collectively viewed each video and timestamped major segments, identified as changes in conversational focus. During this initial pass, team members flagged for further analysis those segments where the students’ conversation appeared to lead to important insights. For the current study, the authors identified instances where the students used their developing statistical knowledge in an effort to understand the output of the devices. These instances were then examined for structural similarities. Through repeated examination and discussion, these similarities were refined into the characteristics presented in this chapter.

CHARACTERISTICS

As noted previously, other publications by the authors have examined the use of wearable devices in an inquiry curriculum (Lee & Drake, 2013b; Lee & DuMont, 2010) and have presented quantitative results demonstrating the effectiveness of this approach (Lee, Drake, & Thayne, 2016; Lee et al., 2015). The characteristics presented in this chapter are intended to provide guidance for helping students view the data from these devices critically and focus their inquiry on the devices themselves.

In examining instances in the video data where students turned their attention to the devices themselves, the authors identified four characteristics in the students’ inquiry: 1) Noticing Discrepancies, 2) Questioning the Devices, 3) Hypothesizing Conditions, and 4) Investigating. Whether due to natural variability or a misunderstanding of what is being measured, discrepant data is a part of measurement. The key for having students critically examine a measuring device (wearable sensor or otherwise) is to provide a means for them to notice that discrepancy. With activity trackers, that noticing may be facilitated by the children’s familiarity with the activities being measured. Inconsistency between a recalled experience and a representation of the resulting data can lead students to question the device and their assumptions about
it. In an effort to resolve their questions, students may hypothesize conditions under which the device might (not) work. Small-group and whole-class discussions are useful for helping students refine these hypotheses into a testable form. Once the students have a testable hypothesis, they can begin an investigation. This investigation includes defining the data collection methods, performing the collection, and then analyzing a visualization of the resulting data. Recollection of the data collection activities can play as important a role as statistical methods in students’ analysis of this data. This analysis may result in further refinement of the conditions and another round of investigation or a resolution of the students’ initial questions.

It should be noted that the goal of these investigations in the current study is an acceptable resolution to the students’ questions based on the data rather than achieving a specific solution. The analysis is often “quick and dirty,” as the students may be learning the processes of inquiry and data analysis at the same time.

EXAMPLES

The examples that follow seek to show how these characteristics operate in practice and explicate their ties to the NGSS Scientific Practices. These examples were chosen because they demonstrate different approaches to data collection and visualization. The first example comes from the first year of the project. During this year, the students used aggregated data across the entire class to examine the accuracy of the trackers. So that students could see their individual contributions to the aggregate display, the authors chose to use a decidedly low-tech solution to data visualization—sticky notes on butcher paper. In the second example, data from the students’ investigations are represented and displayed to the class using the Tinkerplots data visualization software (Konold & Miller, 2005).

Classroom Example 1: Are Activity Trackers Accurate?

At the beginning of this project, activity tracking was a new phenomenon. The public was just becoming aware of activity trackers and possibilities they offered. Prior to this, physical activity tracking was limited to specific episodes (e.g., “going for a run”). While tracking in this manner was embraced by the athletic community (Lee & Drake, 2013a; Lee & DuMont, 2010), the “always on” tracking of routine activity represented a fundamental shift in what activities were trackable and who would be doing the tracking.

In the year prior to this case, the authors had used activity trackers for small-scale studies (e.g., Lee & Drake, 2013b) to identify avenues for their use in classroom
activities. At the time of this case, the researchers were working to develop and implement a unit with a fifth-grade class twice a week from September through December. The curriculum used was designed to promote student-led inquiry, with teachers and researchers encouraging the students to focus on questions that could be answered statistically using data from the trackers.

**Introduction to Activity Tracking**

The unit began with a device exploration period during which the students each wore an activity tracker and moved around the school to see how they worked. After this free-exploration time, the class assembled in small groups to share what they noticed about the devices and come up with a shared idea of how well they worked. The discussion proceeded with the students offering their assessment of how (well) the trackers worked and why they came to that conclusion. The students immediately began questioning how accurately the devices could count steps.

**Jocelyn:** Mine felt inaccurate because when I just took 90 steps, it took me up to about 200 steps and I was very confused.

**Houston:** Mine was perfectly accurate. When I took 10 steps, it added 10 steps...

**Mischa:** Like Jocelyn said, when I walked to the library, it said I took 200 steps, but I actually took a lot less than that.

**Students:** [Chorus of agreement]

In this excerpt, the authors note evidence of both *noticing discrepancies* and *questioning the device*. Each of the students cited their own performance (“I just took 90 steps”) and the on-device count (“it took me up to about 200 steps”) as evidence. The Fitbit Ultras used by these students had an on-device LED display that showed the current daily step count. The students frequently checked this display against their own step counts. Jocelyn and Mischa both recognized discrepancies in these comparisons. Although Houston claimed perfect agreement in his exploration, this claim was largely ignored. Conversely, the chorus of agreement following Mischa’s statement indicated that many of the other students had noticed discrepancies during their trials as well.

Beyond noting discrepant data, the students started questioning the devices using these discrepancies as evidence. In positing that the devices were inaccurate, the students were stating what would become the driving problem for the remainder of the inquiry activity (NGSS Practice #1: Asking questions and stating problems).
Making Hypotheses Public

Following a second free-investigation period focused on the accuracy of the devices, the class reconvened to discuss their findings. The goal of this discussion was for the students to create a shared understanding of observations they made or questions that could be answered using the trackers. These observations and questions were captured on a large sheet of poster paper hung at the front of the room. This list remained visible throughout the remainder of the unit. The segment that follows shows how the language for each item was discussed and refined.

Clara: There could be a little person who takes really small steps and it’s hard for the Fitbit to know if you just took a step or not or there could be a really big tall person who takes big obvious steps and the Fitbit can process it faster.

Erin: I think what she basically said is it depends on the person and the place it’s put and the Fitbit.

While the students could still advance any hypotheses they chose, the class seemed to focus on hypothesizing conditions that might affect the accuracy of the devices. Clara’s suggestion that accuracy depends on the size of the user made sense within the context of the class. It allowed for the devices to be accurate for use by adults (the students knew that their teacher was using an activity tracker) while allowing for perceived inaccuracy in student measurements.

Erin’s “clarification” advanced another condition—that accuracy depended on “the place it’s put,” referring to where the device was worn (i.e., trackers may be more accurate when worn at the waist than on a lapel).

The conditions offered by students can serve to further define the main problem and focus the inquiry. These conditions may also identify new problems. In either case, giving students the opportunity to express conditions under which the device might (not) work, gives the students experience defining investigable problems (NGSS Practice #1). In the case of Clara’s hypothesis, the class took up a version of it in a subsequent investigation.

Through this and the previous class discussions, the students began to flesh out a problem definition—that the activity trackers were inaccurate. In identifying and shaping this problem, the students leveraged their familiarity with the activities about which they were collecting data (i.e., walking, running, and jumping).

Planning the Investigation

After completing the class hypothesis list, the students elected to continue trying to determine whether the activity trackers were accurate enough to continue using them.
for further inquiry investigations. To do this, the students decided to use a sample of seven trackers to try out their test methods. During a whole-class discussion, Erin and James proposed methods that would eventually shape the structure for the test.

**James:** I think there should be one person with seven Fitbits... When they're walking, they count their steps and then they look at all of the Fitbits and see if they all have the same answer.

**Erin:** I think we should have one person focusing on how many steps they take... and then another person just thinking about other things while they're walking.

Planning this investigation (NGSS Practice #3: Planning and carrying out investigations) was a collaborative effort. The suggestions continued to draw on students’ familiarity—direct experience with the devices, earlier discussions, and conditions on the class list created earlier. Investigation planning continued the pattern of refinement begun in previous share-outs. The final-form method that the class agreed to had three roles: a Walker, Counters, and Monitors. The Walker would wear all of the devices. Counters would follow the Walker and silently count the Walker’s steps. Monitors would be assigned to one device each; they would record data on a sticky note—a device identification number, the step count from the Counters, and the starting and final step counts from their device. The Walker and Counters would walk a set distance then return to the classroom. Each Monitor calculated the number of steps recorded by the device (final count minus starting count) and the residual between the device and Counter step counts.

Figure 4. Students testing accuracy of Fitbit Ultra devices by having a single student wear multiple devices, manually counting the student’s steps, and comparing the counts
In planning this investigation, the students collaboratively developed a method that parallels methods used by professional studies looking into the accuracy of these devices (Diaz et al., 2015; Takacs et al., 2014). In these studies, multiple devices are tested per trial, and their output is compared to observer-counted steps. Building off each other’s statements, the students were able to devise a reliable and repeatable method for testing the functionality of the devices. This method resulted in a corpus of directly comparable data points that could then be used as the investigation turned toward representing the data as a means for answering their question.

**Conducting the Investigation**

The students conducted 11 separate trials with varying numbers of activity trackers per trial, resulting in 56 individual sticky-note data points. Following each trial, the trackers placed their sticky notes on a poster with two categories marked: “Not very accurate” and “Pretty accurate.” The difference between the categories was kept intentionally vague to allow the students to decide what constituted “accurate enough” individually. Because this approach would result in a less conclusive representation, the authors and the teacher hoped it would spur the students to desire more formal categorization of the data. Although the “pretty accurate” bin contained a few more sticky notes (see Figure 5), the students did not see this initial

*Figure 5. Detail of a video frame capture showing the first representation of data from the Fitbit accuracy trials. The left column is “Not very accurate;” the right column is “Pretty accurate”.*
representation as providing the conclusive evidence they needed to decide whether the Fitbits were accurate.

To help the students decide for themselves whether the devices were accurate, the authors prepared four posters on which the students could create frequency plots. Each of the four posters divided the data up differently to show the students that different representations can be made of the same data and that the right representation can help answer questions.

Working in small groups, the class created four frequency plots using sticky notes. The finished plots were hung in front of the class for inspection.

In presenting their plots to the class, students from each group reported how the plot convinced them that the trackers were accurate or inaccurate. Each of the presenters said they believed that the devices were accurate, offering as evidence the larger clumps of data in the bins representing lower residual step counts (NGSS Practice #6: Constructing explanations). After each of the groups had presented, the teacher invited Sabine to share her interpretation of the plots. Sabine was invited to share because she had been reluctant to accept anything short of perfect agreement before agreeing that the devices could be accurate. As she stood in front of the plot (Figure 7), Sabine noted the tall peak in the first column and the rapid drop from there to the bins with greater differences. Using similar reasoning, the rest of the class accepted that the devices were accurate enough for continued use.

Before creating the plots (NGSS Practice #4: Analyzing and interpreting data), the students relied on individual trial results when making their claims. After seeing the data in aggregate, the students were able to use a single number (i.e., the mode) to summarize a data distribution, which has been identified as an early stage in learning to statistically analyze and interpret data (Lehrer, Kim, Ayers, & Wilson, 2014).
Example 2: Do Jumps Count as Steps?

The second classroom example comes from an enactment in a sixth-grade classroom. Students in this enactment each wore a Fitbit Flex (Figure 1) throughout the school day. Because the Fitbit Flex does not provide immediate feedback, students had to wait until the next day to see their data. By using the devices in this way, the students could not directly compare step counts from particular activities to output from the devices. As a result, the students generally viewed the trackers as authoritative, a view encouraged by the teacher throughout the unit. However, as in the previous example, the students come to question the abilities of the devices when their output is viewed in light of the students’ own experiences with familiar activities.

This example begins with the students examining three unlabeled sets of data from the previous day’s recess (Figure 8); the activity’s objective was for the class to identify the student represented by each dataset and justify the identification. This activity was intended to encourage students to attend to various features of the data and make connections between distinct features and the students’ recalled activities. With various visual landmarks specified (Moher et al., 2014), students could then
consider how less distinct features mapped onto the data-generating experiences. This data examination activity spurred discussion of what exactly it means to “jump rope” at recess and whether or not this is a worthwhile approximation of what it means to “jump.” This in turn led the students to wonder whether or not the activity tracker recorded jumps as steps, the tracker’s primary data product. To answer this question, the students devised several data collection and analysis activities as they refined their understanding of both the activity of jump rope and how to test the devices.

Interrogating the Device: Initial Talk About Jumping Rope

At the beginning of math class each day, data from a handful of students was projected on the whiteboard in front of the class. The students analyzed the data using a combination of statistical tools and their knowledge of the activities represented in the data (NGSS Practice #4). The analysis involved the students whose data were displayed introducing the activities that produced the data. The whole class would then engage in a discussion to identify features in the representation, compare overlaid data sets, et cetera. The nature of the analysis task changed from day to day in an effort to broaden the students’ analytical repertoire.

The data investigation in this example began with unlabeled recess data from three students (Kelly, Chad, and Jason) projected in front of the class (Figure 8). To introduce the data, these students were asked to share what they did during the morning recess and which data points corresponded to their activities. Kelly claimed the yellow points (white dots in Figure 8) represented her recess because she walked around at the beginning of recess then joined a game of jump rope and did not move much after that. After the other students shared their recess, Nick began the exchange that follows by challenging Kelly’s selection of the yellow points with an alternative interpretation of the data.

Nick: I think Kelly is red ‘cause she said she was just walking around in the beginning, and it’s kind of low. And at the end she said she was jump roping, and jump roping takes a lot of steps, so I think she is red.

Mrs. Bryson: I wonder if she was walking around while she was waiting. I wonder about these zeros though.

Nick: I said red not yellow.

Mrs. Bryson: She said yellow. Right, Kelly? So, you’re thinking she is red instead. Does anyone have anything to add to that?

Researcher: I have another question. Why do you think that jump roping wouldn’t get you very many steps?

Kelly: I was swinging most of the time, but I don’t know
From Wearing to Wondering

Nick’s interpretation accepted Kelly’s general recess narrative—walking around followed by playing jump rope. However, Nick asserted that “jump roping takes a lot of steps,” in opposition to Kelly’s claim of having not moved much after joining the game. This assertion seems to align well with the red data, which have higher step counts at later times. Mrs. Bryson returned to Kelly’s claim and raised a concern that it did not fully account for the zero-valued yellow dots. This statement made explicit an assumption that Kelly had made but not yet stated aloud: playing jump rope involves times when one does not move and thus does not get any steps in a given minute. To clarify Kelly’s reasoning, a researcher asked why she thought that “jump roping wouldn’t get…very many steps?” Kelly responded that she was swinging the jump rope, which does not involve taking steps, although she was unsure if that made a difference.

In this excerpt the discrepancy the students noticed was between the data and students’ expectations. Two views of the nature of jump roping were advanced—one involving stationary periods of time and another that involved continuous bodily movement. With datasets that could potentially be linked with either of these views,
From Wearing to Wondering

there was no easy resolution to this question. The discussion made jump rope a more complicated activity. Further, the assumption that jump rope might produce fewer steps, relative to other activities, was opened for examination by the researcher. In the next excerpt, this assumption was taken up again by the teacher, who helped maintain an ongoing focus for this discussion on the question about whether jumps would be counted as steps.

Asking About the Tracker’s Step Counting

Mrs. Bryson: I have a question about the about the jump rope, this is a Fitbit question, if you are just jumping up and down in the same place…
Kaylie: …does it count?
Mrs. Bryson: Does it count steps?
Alondra: I guess so.
Adam: I would say so because you using, doing exercise. You’re exercising
Carson: I think it thinks that your movement is that you are running around
Mrs. Bryson: Yeah, but the Fitbit is smart, but is it smart enough to do that? Can we find out?

Out of genuine uncertainty, Mrs. Bryson more formally asked the question about whether jumps were counted as steps by the wearable devices. This question changed the focus of the discussion to the device while maintaining jumping as the condition under which the device’s function was being questioned. While Mrs. Bryson first advanced the question, Kaylie’s interjection hints that the question was likely one the students were already anticipating at that moment. Adam volunteered the view that because jump rope was exercise, a fitness-tracking device should track it and make jumping count. Other students seconded the idea that the tracker would interpret jumping as steps. These statements suggested students already suspected that jumps counted, but Mrs. Bryson made an important move next by changing the question from “Does the Fitbit count jumps?” to “Can we find out?” The change of phrasing implied that simply assuming how the devices worked would not be enough and encouraged the students to find a way to answer this question.

Following this, several more suggestions were made about how the class could devise a way to address the initial problem of finding a way to determine if the trackers counted jumps as steps (NGSS Practice #3). This ultimately led to a suggestion from a student that received class support: three students would spend all of that day’s recess playing jump rope, and three other students would be sure to walk for the entire recess. According to their plan, having a walking group and a jumping group would let the class see if the numbers were equal and could tell if jumping was registered as walking (taking steps).
While the teacher initially voiced the question, the students seemed to take it up as their own question to answer. Initially, they offered their assumptions that the device (which they took as presenting the “truth” about an activity) would work. After Mrs. Bryson changed the question to “Can we find out?” the students took it on themselves to plan an investigation (NGSS Practice #3) to use the device’s data to probe a potential limitation of the device.

Reviewing the Data From Walkers and Jumpers

When the class convened the next morning, Mrs. Bryson recapped the class’s plan to determine if jumping was counted as steps by the devices then asked the participating “walkers” and “jumpers” to share what they had done. All six “walker” and “jumper” students reported attempting to do the activities that they were assigned. However, with the exception of Adam, each of these students was eventually absorbed into other existing activities. Some students got “bored” and changed activities, and some played regular games (such as four square) and tried to incorporate walking. For those who did jump rope, the standard play rules had continued to apply. Two of the students stated they had limited time to jump, and the third did not get to jump at all.

Based on how students tried to express that they had at least attempted to walk or jump amidst their recess and how animated students were during the previous day’s discussion, it did not seem that the students were being dismissive of the data collection activity. However, other aspects of recess, such as play, socializing, and

Figure 9. TinkerPlots graph comparing “jumpers” (black dots) and “walkers” (white dots)
Following established playground rules, took precedence. Because the conditions for this test were set explicitly within established recess activities—jumping rope and walking—the influence of established recess order in the data collection should come as no surprise. These conditions harkened back to the initial data analysis task (i.e., identifying data corresponding to a student who jumped rope) that led to the question of whether the devices counted jumps as steps.

When the data were projected (Figure 9), the students could see how these conditions were problematic for answering the question, leading them to discuss the validity of their experiment’s design.

**Evaluating the Design of Their Investigation**

After being presented with the students’ recollections of recess and the accompanying data plot, Adam began the following exchange by questioning how data should be interpreted in light of his own observation of a jumper, Melissa.

**Adam:** I watched Melissa, and I saw her, and she was standing in line for the jump rope a lot.

**Mrs. Bryson:** She said that when she talked about that.

**Kate:** We have to stand in line, because you’re not really...

**Adam:** I know, I’m just saying, I am not blaming you or anything. I am just saying, just that maybe it is not very accurate because the walkers walked the entire break, but the jumpers just jumped some of the break and part of it they were standing in line.

Adam recalled seeing Melissa “standing in line for the jump rope a lot” during recess. Due to the layout of the playground and his own role as a walker, Adam was able to watch what others were doing on the playground when they were supposed to be walking or jumping. That position of observing what happened when data were collected made the regular noticing of what was happening on the school playground a relevant resource for the conversation.

His main point was that there was a discrepancy between the conduct of the investigation (i.e., jumpers spent too much time standing in line) and conditions sufficient to answer the question (“Do activity trackers count jumps as steps?”). This point echoed the conditions that served as the impetus for the investigation, but in the context of Adam’s statement it was important because the nature of the test itself was being questioned again. Adam recognized that the experiment as run did not meet the conditions described in the question—the recess data included periods of standing, jumping, and several other activities. In order to draw any conclusions relating to the main question, the investigation needed to be refined.
Proposing a New Investigation

This concern was recognized by the class and Mrs. Bryson. For several minutes, students offered a number of possible ideas for new data collection. These ideas were vetted for feasibility and utility. In the segment below, Casey suggested the new data collection activity for the class that would then be implemented.

**Casey:** We could do it (jump) for a minute and count how much we do and then we can look at it (data) tomorrow. And we could look at someone who remembers how many they have and they could write it down or something, and we could see how much steps it counted.

**Mrs. Bryson:** Can we do that? We can do that right now, couldn’t we?...

**Jimmy:** Just do it for a minute... Wait until the next minute and then we stop right where the time is and then we just look at [the data].

The students’ plan for this new investigation (NGSS Practice #3) involved all the students jumping for one minute during math class. After getting similarly inconclusive results from looking at data from two recesses, the students decided to perform data collection during class. This choice freed them from the constraints of recess and shortened the time in which the students needed to focus on data collection. Having all the students jump allowed everyone to participate and resulted in a similar number of points to the previous data sets the students had analyzed. The students’ decision to count their own jumps for comparison against the tracker counts parallels the investigation in the previous example, with the students’ self-counted jumps being the authoritative count.

On a signal from Mrs. Bryson, the students all proceeded to jump, with many students counting aloud. When the minute ended, the students excitedly began calling out their numbers of jumps; Mrs. Bryson told the students to quickly write the numbers down so they could compare the next day after the data from the Fitbit devices had been synched and the minute of jumping had been made visible.

Reviewing Data From the New Investigation

The next and final day of this unexpected jumping investigation began with Mrs. Bryson saying that she had access to all of the students’ counted steps during the minute of jumping from the previous day. She called on various students to share what number they had counted and recorded themselves and then reported back the number that the Fitbit devices had counted. The numbers that were obtained by the Fitbit devices generally within about 15 jumps, with a few students having numbers that differed by several dozen. Following this, Mrs. Bryson asked how
many students had their own count as higher than what was recorded on their Fitbit device (10 students), how many had an exact match (1 student), and how many had counts below the Fitbits (8 students). The class then examined the distribution of steps recorded during the minute of jumping and discussed what were the various measures of center for it (NGSS Practice #4).

Eventually, Mrs. Bryson suggested to the class that even though the numbers were not exact, it did appear that many of the Fitbit observations of steps were close to human counts. For the purposes of the class, that meant the Fitbit would usually count jumps as steps. Even though that was established as understood and accepted knowledge for the class, some students still pushed for other possible variations on their test in an attempt to mitigate sources of error. Mrs. Bryson noted the cleverness of these suggestions, but in the interest of getting back onto the unit as planned, steered conversation to other topics related to data and distribution.

SOLUTIONS AND RECOMMENDATIONS

The examples above show the students from two elementary school classes participating in an inquiry-based statistics unit that featured wearable activity trackers to produce data for analysis. The authors’ goal for this unit was to see how students could leverage their familiarity with their own activities as a tool for analyzing and interpreting displays of activity data. The thought being that students would develop a deeper understanding of both their own activities and the statistics content. On occasion, however, the students turned the tracker’s data back on itself, making it the object of inquiry. In these instances, the authors noted certain characteristics to the students’ inquiry. The students 1) noticed a discrepancy in the data (either relative to their personal experience or to their expectations), 2) questioned the device, 3) hypothesized conditions which might explain the discrepancy, and 4) planned and carried out an investigation.

Students’ ability to notice and call out discrepancies in the data is key in this process. Recognizing irregularities requires a degree of familiarity with both the data and the processes by which those data were captured. Because wearables capture data related to the wearer’s mundane experience, familiarity with the data collection activities is assumed to be built in. This familiarity brings with it expectations of how the activity is supposed to proceed, which students do indeed bring to bear on data interpretation tasks (Lee et al., 2015). However, students’ familiarity with representations of the resulting data should not be taken as given (Ching & Schaefer, 2014). To help students develop the same degree of familiarity with data that they have with their own activities, they need to be provided frequent opportunities to
From Wearing to Wondering

inspect and interact with that data (Lee et al., 2016). In the curriculum used in this study, the students viewed and publically interpreted their data on a daily basis.

Throughout the students’ analysis activities in this study, the students’ recollections of their experiences were treated as valuable sources of information. The teachers encouraged students to share this knowledge, particularly as it related to the data being examined. Establishing classroom norms requiring students to ground their explanations in data is a powerful tool in inquiry (Ryu & Sandoval, 2012).

The students in the examples above were able to resolve their questions about the devices to their satisfaction. Along the way, they were able to gain valuable experience with many of the scientific practices targeted by the Next Generation Science Standards (Achieve, Inc., 2013). It should be noted that the investigations and analyses the students engaged in and their eventual resolutions are all of the “rough and ready” variety. The driving questions were binary, yes-or-no questions—Are the trackers accurate? and Do they count jumps as steps? The resolutions really only hold within the context of the classroom and for the students involved. That is okay. In each of these examples, the investigation was the students’ first experience with inquiry learning, and they were only just beginning to learn much of the statistical content for their grade level (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). These early, small-scale experiences with inquiry can give students the confidence necessary to continue to more advanced investigations and analysis techniques.

FUTURE RESEARCH DIRECTIONS

The research that has been done with wearable activity tracking technologies has shown the promise of these devices in inquiry-based STEM education. However, given their relative newness, much may still be done to better understand what these devices may afford for education. The authors note several avenues that they feel remain to be explored. Among these are the following:

• Expanding the Devices’ Uses With Students of Different Ages: The current study worked with fifth and sixth grade students. Small scale studies (e.g., Lee & Briggs, 2014; Lee & DuMont, 2010) and afterschool clubs (e.g., Ching & Schaefer, 2014) have begun to examine how high school students and undergraduates (Thayne & Lee, 2016) might use the data from these devices. Further work remains to be done to understand how other grades might use these devices and how those uses might be leveraged for productive STEM inquiry.
Expanding the Breadth of STEM Subjects: As noted above, existing research has incorporated wearables into a broad range of STEM content areas. It is not hard to imagine these devices contributing to even more STEM topics. The expanding capabilities of these devices immediately lend themselves to uses in Physics and Biology. With a little creativity, even broader applications are possible.

Examining Equity: Wearable technologies may have the potential to democratize access to a broad range of health-related data and the greater personal insight that comes through its analysis. However, use of wearables “in the wild” is dominated by wealthy, middle-aged white males. The authors have done some initial research to examine how underrepresented populations (do not) use these devices (Lee & Briggs, 2014), but more remains to be done. Only by understanding the motivations behind their (dis)use can they be used in a culturally-responsive, and truly productive, manner.

CONCLUSION

Wearable technologies are becoming more prevalent in everyday life. Using these devices as tools for inquiry holds many possibilities for understanding a variety of STEM topics (e.g., Moher et al., 2014; Resnick et al., 2000). The current study used wearable activity trackers as part of an elementary school statistics curriculum. While the goal of this curriculum is for students to use the data from these trackers as a tool for learning statistical content while examining their own activities, the students occasionally turned those data back on the devices in an effort to answer questions they had about the devices. This chapter presented characteristics of these instances that resulted in productive engagement in inquiry. These characteristics are that the students 1) noticed a discrepancy, 2) used that discrepancy to question the device, 3) hypothesized conditions under which the device might (not) work, and 4) planned and carried out an investigation to test their hypotheses and resolve their questions. These characteristics were then situated in two examples from different years of the current study, each using different activity trackers and means of representing the data.

In the above examples, the students were familiar with the physical activities that had generated the data, both as active participants and as observers monitoring the activities. Discrepancies between students’ expectations for these activities and the data provided by the trackers provided the impetus for the students to question the devices. Through whole-class discussion and with the teachers’ support, the students were able to develop tests that they could carry out using the trackers and would
resolve their questions. Students leveraged their familiarity with their activities to even question the conditions of their own investigations.

Through these inquiry experiences, the students were able to demonstrate many of the science practices in the Next Generation Science Standards (Achieve, Inc., 2013). They asked questions that could be answered using scientific methods (NGSS Practice #1). They collaboratively planned and carried out investigations (NGSS Practice #3). Using statistical methods and their own experience conducting the investigations, they were able to visualize, analyze, and interpret the resulting data (NGSS Practice #4). In the second example, they noticed that their data collection activity would not answer their question and iterated through these practices again.

As researchers and educators work to integrate advancing information and communications technologies into STEM classrooms and reignite interest in STEM topics among students, wearable activity tracking devices represent one option. By enabling students to investigate data from familiar and routine activities, activity trackers can be used to foster inquiry focusing not only on those activities but also on device itself.

REFERENCES


From Wearing to Wondering


President’s Council of Advisors on Science and Technology. (2010). Prepare and inspire: K-12 education in science, technology, engineering, and math (STEM) for America’s future. Retrieved from http://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports


**KEY TERMS AND DEFINITIONS**

**Activity Tracker**: A subcategory of wearable technologies with a feature set that emphasizes collecting of activity data such as steps taken, calories burned, and distance traveled.

**Fitbit**: A company specializing in the production and marketing of a line of wearable activity trackers.

**Personally-Relevant Data**: Data collected about activities or topics of particular importance to the individual(s) interpreting the data.

**Tinkerplots**: Software for supporting data visualization and analysis in elementary and middle grades, developed by a team from University of Massachusetts lead by Konold and Miller.
Typicality: A rough approximation of where most of the data are in a given distribution. Typicality was used as an entry point to more canonical measures of center such as mean and median.

Variability: A measure of how much the points in a dataset differ from each other and from what is typical or average, can be quantified as standard deviation, range, variance, etc.

Wearable Technology: A category of technologies that are worn as part of the technology’s standard use case. These devices often include motion sensors for capturing data relating to the wearer’s activities.