Opportunistic Uses of the Traditional School Day Through Student Examination of Fitbit Activity Tracker Data

Victor R. Lee, Joel R. Drake, Ryan Cain, Jeffrey Thayne

Utah State University

Logan, Utah, USA

victor.lee@usu.edu, jrichdrake@gmail.com, ryan.cain@aggiemail.usu.edu,
jeffrey.thayne@aggiemail.usu.edu

ABSTRACT

In large part due to the highly prescribed nature of the typical school day for children, efforts to design new interactions with technology have often focused on less-structured after-school clubs and other out-of-school environments. We argue that while the school day imposes serious restrictions, school routines can and should be opportunistically leveraged by designers and by youth. Specifically, wearable activity tracking devices open some new avenues for opportunistic collection of and reflection on data from the school day. To demonstrate this, we present two cases from an elementary statistics classroom unit we designed that intentionally integrated wearable activity trackers and child-created data visualizations. The first case involves a group of students comparing favored recess activities to determine which was more physically demanding. The second case is of a student who took advantage of her knowledge of teachers' school day routines to test the reliability of a Fitbit activity tracker against a commercial mobile app.

Categories and Subject Descriptors

K.3.1 [**Computer Uses in Education**]: Computer assisted instruction (CAI).

General Terms

Measurement, Human Factors.

Keywords

Elementary school, wearable technology, activity trackers, design research, self-data, Fitbit

1. INTRODUCTION

In many ways, school-aged children in the United States have little freedom to design their days. Parents and other adult authority figures often dictate what children are allowed to do and when they can do it. As a result, children are limited in where they can choose to go, how they choose to get there, and how much time they can spend immersed in a given activity. The school day is a canonical example of this. When children are at school,

behavioral routines are already well established. Children are to come in, sit down, and do assigned work. They are to move to different rooms only at designated times and on designated days. Even children's means of getting to school is often beyond their ability to dictate. For instance, while playing a fitness learning game where increased physical activity yielded more resources in a virtual game world, many children began walking to school. However, many students felt disadvantaged because they did not have the option to choose to walk to school, because of parents' rules, geographic location, or other constraints [4].

It is then not surprising that many efforts to design meaningful and engaging technology interactions that go beyond "sitting in the computer lab" involve non-school settings, such as after-school clubs, museums, or online environments [11,22]. In those spaces, children are thought to have more flexibility with respect to what activities they can pursue. The standard time limits typically established by school bell schedules are no longer imposed. Identities and roles taken by youth and adults are also far more fluid [3,10]. When a comparison is drawn between school and non-school settings, schools appear to be far more constrained and less attractive spaces for creating innovative child activities and interactions.

However, the contention of this paper is that while there are admittedly constraints associated with children's days—and specifically from compulsory attendance at formal schools—there are also new and creative ways to leverage routine school activities for novel purposes. For those who want to alter the kinds of interactions with and among children who spend a substantial amount of time in traditional schools, we should be open to exploring ways for designers and for children to take advantage of existing school routines.
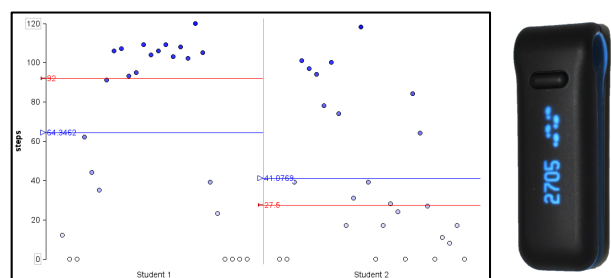
Figure 1. Tinkerplots window displaying data for two students collected during recess (left) and Fitbit Ultra physical activity tracker (right).

To illustrate how this is possible, we present two cases that come out of a multi-year design-based educational research [27] project with fifth-grade classrooms in Northern Utah. As a research and design team, we have committed multiple years to the

development of classroom lessons and activities incorporating specifically selected technologies (i.e., wearable Fitbit activity trackers, a custom web tool for data extraction, and *TinkerPlots [13]* data visualization software, see Figure 1) with the hope of helping students to engage in practices of elementary statistical analysis. The two cases show different ways in which school day routines can be opportunistically leveraged by designers and by children. The first case involves students challenging and then re-examining the amount of physical activity involved in favored recess games. The second involves a student collecting and inspecting data from two teachers' routine afternoon movements to compare the reliability of a wearable tracking device and a mobile app. By presenting the two cases in this paper, we seek to demonstrate that, in spite of the seemingly rigid constraints on children's school day activities, there are productive ways to be opportunistic within the structure of a traditional school day and in service of encouraging children's meaningful inquiry and investigation experiences.

2. RELATED WORK

In educational design research, there have been some notable past efforts to harness aspects of school routines and norms in ways that are consequential for instruction. One noteworthy example comes from the Cheche Konnen project at TERC, in which immigrant students who had prior opinions about the quality of water from water fountains on different floors in their school building (which influenced which fountains students were willing to use). These opinions became subject to a series of student-led tests of water quality and student taste preference [26]. Other efforts include units designed to help students investigate the nutritional quality of their school lunch [12], to photo-document and reflect on typical activities in a classroom [5], or to examine overall changes in recess activity over the course of a week [18].

In recent years, the availability of new technologies has enabled yet more opportunities for students to inspect aspects of their school day experiences. For instance, growing commercial interest in wearable activity tracking devices has reduced the cost of these devices to the point that they are viable options for turning portions of the school day into objects of inspection [19]. Although these devices are typically marketed toward adults and individuals concerned with improving their health or wellness, students can and are beginning to use them to capture personally-relevant records of their daily activities [17].

At their core, these devices are simply sensors (accelerometers, to be precise) with proprietary algorithms that enable them to make a reasonable approximation of some quantified physical activity. Most often, this includes steps taken, calories burned, and distance traveled. Newer, wristband-based devices can also record heart rate, elevation gain, and geo-location. These features are many of the same ones being touted with the rise of "smartwatches." While there actually has been a long tradition of sensor use in education and in schools, such as was the case with microcomputer-based labs and with probeware [28], wearable activity trackers are sensor suites that offer key and unique affordances [19].

The primary affordance to using *wearable* activity trackers in schools, in comparison to older standalone sensor technologies, is that they will act in the background of an ongoing student activity. Whereas probeware required the user to be actively engaged and focused on data collection activities, a wearable activity tracker will passively gather information about the wearer's movement, caloric expenditure, or elevation gain without requiring that individual's attention. This means a child can gather activity data

without disrupting their regular routine, and then subsequently examine and reflect upon these data [23]. Students can encounter and engage with new representations of information on topics and experiences of which they are already highly familiar and are arguably experts [8].

In some earlier work, this overarching approach has shown promise for supporting student learning. We have seen evidence of significantly greater gains in certain elementary statistical topics over traditional instruction [20] and students building on prior recollection of their subjective experience of activities to develop new and more appropriate visual analysis strategies [19].

3. RESEARCH AND DESIGN CONTEXT

For multiple years, our team has been engaged in a program of design-based educational research [27] with two fifth-grade classrooms that involve wearable activity trackers and students becoming acquainted with practices of data representation and analysis. This program has involved partnerships with a public elementary school and its teaching staff to design learning interactions and materials and support implementation of data-centered activities in the classroom. The school is in a small city in Northern Utah and has over 30% of its student population receiving free or reduced lunch, comparable to other schools in the area. As is the case for the geographic area, the school population is predominantly white.

Over multiple design iterations, our team created and refined a set of lessons, inspired in part by the Quantified Self movement and other technology-supported tracking practices [16], that encouraged students to explore wearable device capabilities and prepared them to participate in a formal examination and critique of some traditional methods of representing data, including time-ordered displays and histograms. These preparatory lessons included variability and distribution as they relate to measurement. Variability and distribution have been recently identified as both attainable and essential topics for elementary students to master if they are to make meaningful progress in learning statistical content [14,21,25]. In an earlier year, for instance, we found that students were initially skeptical of the accuracy of step data obtained from Fitbit Ultra clip-on devices. After some systematic testing and exploration of variation, students were ultimately convinced that the wearable devices were accurate enough for their measurement purposes when allowing for a normal pattern of variation (see Figure 2).

Figure 2. Students testing accuracy of Fitbit Ultra devices by having a single student wear multiple devices, manually counting the students' steps, and comparing the counts.

In a later project year, we realized that while students were able to engage individually with their own data, they often lacked common focus and language for regularities across different individuals' datasets. Thus, in the more recent design instantiation from which our case studies originate, we opted to have students examine one individual's activity data as a class in depth initially.

This was an appropriate move because, as pointed out in other research, students can and do exhibit difficulty or initial resistance to canonical dashboard activity data displays (such as the one provided at fitbit.com, Figure 3) [4].

Figure 3. Activity display and aggregate steps tiles from the Fitbit Dashboard, fitbit.com

One opportunistic solution we devised was the implementation of what we called "Fitbit starters." These were short data exploration episodes that took place each day during what was known in the partnering school as "morning meeting." At the beginning of each day, the students gathered for 20-30 minutes on a rug in a corner of the room for a community discussion facilitated by the teacher. These discussions covered topics including recent concerns at the school (e.g., bullying, or helping students dealing with personal grief) and plans for the week or later in the school year (e.g., assemblies, upcoming field trips, or fundraisers). Students often took turns during morning meeting sharing experiences or personal observations that were important to them, and others were given time to comment. Sometimes conversations would even involve casual discussion of what students had done during the previous day or weekend.

When our designed unit began in the current year, the teachers participating in our study allotted some of the morning meeting time to be spent with a student showing and discussing what activity data they had collected the previous day. Having these conversations was a natural extension of the morning meeting routine. As far as we were concerned as educational designers, this public display of data was also intended to familiarize all the students with select canonical data displays on the Fitbit dashboard and make their interpretive strategies communal. Students would state what they noticed and point at key features so that other students would know what was being referenced and interpreted at a given time. Through observing Fitbit starters, we noted that students often engaged in a kind of joint reconstruction of the previous school day's events as they sought to account for a given individual's increases and decreases in activity [15].

After a couple of weeks of Fitbit Starter activities and preparatory lessons related to recognizing variation in activity measurement, students worked individually or in small groups on custom projects using the technologies available to them. Researchers from our team were on hand to help students with data downloading and cleaning so that the students could devote most of their attention to posing questions and creating and looking at visual representations of their data. During these implementations, researchers from our team video recorded each day's lesson. Additionally, we used multiple forms of pre- and post-assessment, specifically written tests and semi-structured interviews. These all also involved a staggered-implementation control classroom.

For the current report, we generated two cases by combing through 33 hours of video footage and identified and marked all instances involving a selected group of students speaking. We also drew from explicit student statements made in video-recorded post interviews where they explained their wearable device activity data project to a member of the research team. These

interviews lasted from 30 minutes to an hour each. Together, these and the physical activity data that students gathered for their investigations were all reviewed to produce the following two narrative cases.

The first case is one in which favored recess activities became an object for student inspection and dispute. The second is a student-conceptualized reliability comparison of the Fitbit Ultra tracker and an iPhone app involving teachers who had fairly routine schedules. Following discussion of these two cases, we present a comparison of student scores from the class that housed these case students against the aforementioned control classroom, comparing what was understood at the same moment in time between the two classes. We also will show what gains were made by the control class before and after they separately completed similar physical activity inspection lessons and custom projects.

4. CASE 1: COMPARING TWO RECESS ACTIVITIES

4.1 Initial Motivation

The first case focuses on three students: Geoff, Lauren, and Neill (note: all participant names in this report are pseudonyms). The motivation for a project they pursued originated during a morning meeting Fitbit starter when they discussed a student's data from a previous day. During that day's discussion of data from that morning's focal student, Eric, the class noticed two large spikes in the data, with one larger than the other. When the teacher asked Eric to describe what he saw in the data display, he immediately noted the first spike and attributed it to being a period of much higher intensity of activity. When asked by his classmates what he was doing during this high-intensity activity, Eric said he was participating in recess. When pressed for more details, he said he had been playing football with his friends.

As Eric proceeded to narrate the rest of his day, he commented that there was another "spike" around midday that corresponded to lunchtime, which also included a much longer recess. This second spike was not as tall as the first. Again, other children in the class had asked what he had done the previous day during lunch recess. He responded that he had played basketball, rather than football again. As is common in elementary schools, some amount of posturing and identification with activities immediately took place. Students (largely males) began speaking out of turn and taunting that football was a more demanding activity than basketball. This quickly escalated to a claim that football was the most demanding of all possible recess activities, a claim which many of the boys in the class verbally accepted.
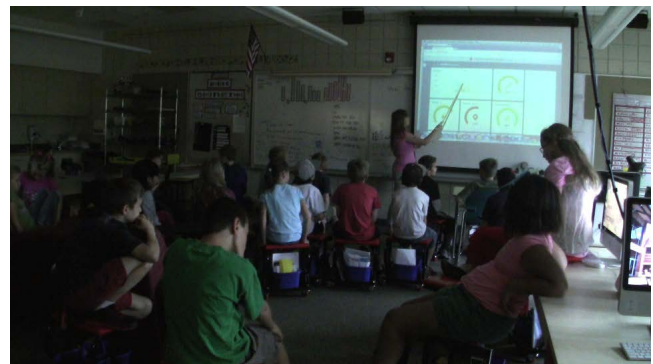
Figure 4. Class participating in a "Fitbit starter" discussing Eric's data.

The importance of this incident is that it struck an immediate chord with a group of students who did not play football at recess (or basketball). Most of the students who played football in that fifth grade class were the more athletically inclined males. Only on occasion did we observe a girl participating in the recess football games. Among the more athletically inclined girls, basketball and soccer were preferred recess pastimes that still required an acceptable level of exertion. Soccer and football were both played on a large field adjacent to the standard school blacktop area. The activity of the day often depended on which group of kids had gotten to the field with proper equipment first.

Figure 5. Class playing soccer for data collection purposes.

Despite the vocal support for football, Geoff, Lauren, and Neill, remained skeptical that football could be the most physically active (i.e., require more steps) recess sport, particularly in comparison to soccer—their preferred sport. Football, they noted, involved periods of time where several children stood motionless in a line, while soccer had no such requirement. When the time came to pursue a custom project, it was imperative to them to prove that football was *not* the most active sport by showing instead that soccer was more active than football. They just needed to get data to support that position.

4.2 Resulting Inquiry

Geoff, Lauren, and Neill convinced the class to participate in a head-to-head comparison. They thought that football and soccer would make for an interesting comparison, since they are both active, and it is not intuitively clear which is *more* active based on personal experience alone, as Geoff shared in a post-interview:

> Geoff: **Football and soccer are such active sports, when you come back in the room everyone has red faces and head to the drinking fountain**, Four Square is just a recess game, you don't move as much in Four Square, a lot of other people don't even know what Four Square is. Most everybody knows what football and soccer are, it's nationwide [sic] known.

Because the entire class was already wearing Fitbit activity trackers, data collection did not involve any additional work beyond having everyone in class play the same sport for a 20-minute block of time. The class data collection was completed over two days, with kids being asked to play football on one day and soccer on the second. Once data collection was complete, our research team downloaded the activity dataset (1120 data points) using a custom PHP-based online tool we created (http://ecds.ed.usu.edu/fitbit), and provided it to the students for inspection. However, when they saw the data, they observed there were a number of extraordinarily low values. These, they realized,

were from kids who, for various reasons, did not really participate in one or the other sport.

> Geoff: We didn't use everyone's data because some of them weren't even playing the game, some people were just standing there, people were walking around randomly, some people were not even playing the game so **we chose people who we knew were playing because we had seen them play**… We wanted to include the people who were actually participating otherwise it would be false data points. It would just be data points that would make soccer better because some people didn't want to play football.

One thing to note in the data inquiry is that the students confronted real issues related to outliers and data cleaning. They had to make decisions about what data were worth including and what would make for a fair comparison. These are rather sophisticated considerations for a fifth-grade student, but because they were there and could see how people were participating in the activities in person and had developed an understanding of how recess activity data typically looked from participating in Fitbit starters, they felt able to make appropriate decisions.

> Geoff: We knew these people were not participating; it was going to be adding more weight—false data. I think of data as weight. **If we're adding people who just played soccer and not football, and it's just adding dead weight**.

The three students eventually decided that, in order to be fair, they needed to include only those students who legitimately participated in both days. So, students who were absent one day or who were visibly not participating in one sport but were participating in the other were excluded from their dataset.

Once they compiled the data that met their criteria, they set about analyzing their data using the *TinkerPlots* data visualization software. After several attempts to organize their data, they created the following visualization (see Figure 6).
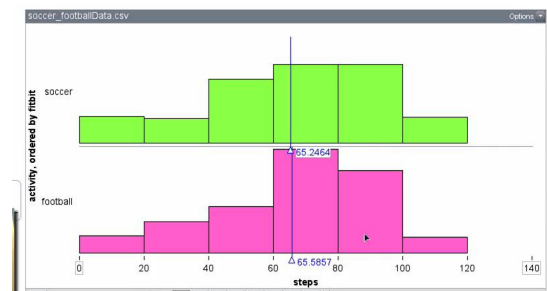
Figure 6. Data display prepared by Geoff, Lauren, and Neill comparing numbers of steps taken per minute while playing soccer or football.

Upon visual appraisal, the three students decided that the shapes of the two distributions were more or less comparable with one another. Based simply on overall shape, there were no major differences of concern. Consistent with the kinds of data they had examined in the preparatory lessons they experienced before, both soccer and football histograms peaked in the same bin and tapered along each side. This meant that there was no easily discernable difference.

Next, the students moved beyond informal visual inference based on distributional shape to look at the mean values for the two sports they had compared. Using the 'show mean' feature built into *TinkerPlots*, they found that soccer produced an average of

65.2 steps per minute and football averaged 65.6 steps per minute. Based on these values, it could have been reasonable to infer that football involved more activity (at least in steps) than soccer. However, they decided that this difference was not entirely meaningful. As Geoff explained:

> Geoff: **It's really not much of a difference,** it looks like a difference at first but when you really look at it, it is not much of a difference it looks like a step, but **it is not even a step.**

A logical follow up analysis could be to test for statistical significance, but this was more advanced than what we had expected the students to do. Instead, the students drew on their own prior knowledge, having routinely participated in these sports and seeing representative data in morning meeting Fitbit starters, to determine that an average of a fraction of a step more per minute (0.4 steps/minute) was not a meaningful difference for their purposes. Furthermore, Geoff added the following:

> Geoff: Once you think about it and look at these you think one step, who cares if it takes one more step. One step is not even going to burn ¼ of a calorie, it won't even make you more fit so you might as well do one or the other because they are practically the same. **If it took 100 more steps then possibly you would choose one [game] or another, but if it takes ½ a step…then it doesn't matter there is not anything extra.** The game is fun, but which one of these you think is fun is the one you should be playing because they have the exact same amount of steps taken.

In the end, leveraging their knowledge of and experience with routine recess activities enabled these students to conclude that soccer and football could be viewed as requiring more or less the same level of activity as one another for the students in their class. For the time being, an athletic truce had been drawn.

4.3 Things learned and demonstrated

Through this set of experiences, the students learned and demonstrated a number of things. First, they noticed that recess varied and discussed how it varied, as shown in the initial morning meeting when football was believed to be the more athletically active recess game. This belief and a desire to disprove it motivated further inquiry.

Geoff, Lauren, and Neill also learned about real issues involved with data cleaning and dealing with extreme outliers. They had points that did not conform to what the rest of their data were showing, and they used their own recall of the previous days' events to make determinations about what data to keep and what data to exclude. They were also demonstrating a sense of what made for a fair comparison and only included individuals who had participated in both sports so as to not overly 'weight' the numbers on one sport.

In addition, the students also demonstrated some sophistication in the methods they used for looking at larger sets of data. They examined the distributional shape and location of histograms for major differences and used those to make inferences. In this particular situation, informal inference could not provide a satisfactory answer, so the students used the mean for comparison. But even when they used a single numerical value (note: they did not explore nor learn about standard deviation), they developed a meaningful, practical interpretation; because the means differed by less than a full step and partial steps did not warrant concern in this context, the difference in activity demands between soccer and football was not "significant."

4.4 Opportunistic Use of the School Day

This case highlights two ways in which our interaction design used the school day structure opportunistically. First, we incorporated data analysis activities (i.e., Fitbit Starters) into the school's morning meeting. Morning meeting was already a time set aside for students to safely air their personal sentiments publicly. Incorporating the Fitbit starter activities into the morning meeting allowed students to apply those norms to interpreting and discussing data displays.

The second way our design opportunistically leveraged school routines was the deliberate examination of data from recess activities. Recess holds a special place in the school day as a small window of time during which students are allowed to pursue activities of their own choosing. These activities and the freedom they represent were of great personal value to individual students. Using wearable fitness devices, these activities also became objects of both inspection and contention. Although we previously pointed out that students do not often have the opportunity to design their days, recess still maintains a special status as period in which the children make the rules and choose what to do. This freedom gave them a time to focus upon and find some variation from what other students might have.

In this case, the students opportunistically leveraged their familiarity with and affinity for particular recess activities in the formation of their research question. While discussing a classmate's dataset (i.e., Eric's data), members of the class began to draw inferences about the relative activity levels of various recess activities (e.g, football was most demanding). Hearing these sentiments but lacking supporting quantitative data, Geoff, Lauren, and Neill drew on their experience with recess activities and their preference for soccer to form their research question: Which sport demands the most of students at recess? Thus, combining data analysis activities with an open-forum morning meeting provided valuable opportunities to stimulate questions of interest to the students. It also enabled the students to draw on their own specific recollections of who actually played both sports so that they could make a fair comparison.

5. CASE 2: COMPARING DEVICE RELIABILITY DURING TEACHER AFTERNOON ROUTINES

5.1 Initial Motivation

The second case is of a single student, Emily. During our study, we tried to capture interactions between students as they grappled with questions of data analysis. This typically favored student groups who were vocal and could productively engage with one another. Because Emily worked on her own, we initially avoided recording her. During a routine check-in to make sure she was making reasonable progress, the second author recognized that Emily was engaging with her data in interesting and at times, quite sophisticated ways. After that, we assigned a researcher to join Emily, primarily to give her someone with whom to talk and to make her thinking visible for us. We present her as a case here to demonstrate how a student can opportunistically leverage the school day structure of teachers and staff rather than that experienced by students, as was shown in the previous case.

The focal student, Emily, was in her first year at our partnering school. She was soft-spoken, attentive during lessons, and naturally curious—in many ways an ideal student. Because of her demeanor, she often seemed more comfortable interacting with some of the teachers than with other students in her class, which

partially explains her decision to work alone. On the annual fifth grade outdoor science education field trip [15], Emily met Mrs. Thompson, one of the third grade teachers who was serving as a chaperone. Following the field trip, Emily volunteered in Mrs. Thompson's classroom on afternoons when she finished her assignments early, which happened quite frequently. During her volunteer time, Emily became familiar with the rhythms and routines in Mrs. Thompson's class, including Mrs. Thompson's own relative activity in the afternoons.

When the time came to select a subject for investigation, Emily decided to compare the "accuracy" of the Fitbit Ultra tracker and an iOS pedometer app. (In actuality, it was more a test of reliability of the two products.) She drew on inspiration from both her home and school experiences in developing this idea. In describing one inspiration for her comparison, she said the following during her post-interview:

> Emily: I was looking at my [older] brother's iPod and saw a [step tracker app]… in his "useless" folder and I thought… **my Fitbit does the same thing. Let's compare the accuracy of it**.

Despite the iOS app being labeled "useless" by her brother, Emily determined to systematically test for herself how the "accuracy" of a step counting app compared with a Fitbit. Rather than using the "useless" app, Emily used Mrs. Thompson's favorite step tracking app because she had, in Emily's words, "tons of pedometers on her iPhone," a fact she became aware of while volunteering in Mrs. Thompson's class during the afternoons.

Because students were not permitted to use an iPod/iPhone during school, Emily could not participate in her own study. Rather, she enlisted two adults with whom she interacted regularly, and whose activities she was able to observe—Mrs. Thompson and a student teacher from her own classroom with whom she had also grown friendly. Thus, with the two teachers available, she was ready to begin a multiday comparison.

5.2 Resulting Inquiry

Every day for two weeks, Emily left her class at the start of lunch to provide Mrs. Thompson with a Fitbit activity tracker (the student teacher already used one). She would also make sure that each participant started her phone app and reminded each to carry her phone with her as much as possible. Emily then gave them each a sticky note on which she had written the date, each teacher's name, and the starting count for each step-tracking tool they were each to use. Each afternoon, the teachers were to return the sticky notes and Fitbit (for Mrs. Thompson) to Emily with the final numbers for the day. After receiving the sticky notes, Emily entered those into Tinkerplots as data and calculated how many steps had been counted (see Figure 7).

Figure 7. Emily with the day's post-it notes from the two teachers entering data into a table.

Despite having two willing participants, Emily still ran into challenges with data collection and data quality. End-of-day

teacher activities (e.g., supervising the car pool pick-up zone), inadvertently deactivating the app, or even wearing clothes without pockets affected the number of steps counted by the step counting app; without pockets, the teachers had to tuck their iPhones into other articles of clothing, typically a boot. Additionally, the teachers occasionally forgot to record their step counts or would lose the sticky note before they could get it to Emily. Because the Fitbit website reported step counts in 15-minute increments, the teachers were able to provide archived step information from their Fitbits. However, the app only reported total steps, so retrospective data from the app was frequently distorted or not saved and, therefore, unavailable.

After looking over her data set, Emily decided that she could not use days of data where she had information from one device but not the other because the actual measurements from the same events were not present. So, while she had her two teachers obtain data over multiple weeks, she did not have a full set of data points to reflect that.

As she analyzed the data, she took personal note about unusual activities. For instance, she noted if the teacher had to put the iPhone in a boot rather than in a pocket. She noted days that had mismatched data. She also noted one day in particular when she heard on the intercom that Mrs. Thompson was being summoned to go downstairs to the main office unexpectedly for an emergency meeting in the afternoon, which she knew was atypical. Not all of these details were maintained carefully nor were they included in her resulting data representations, but she did factor these things into her analysis.
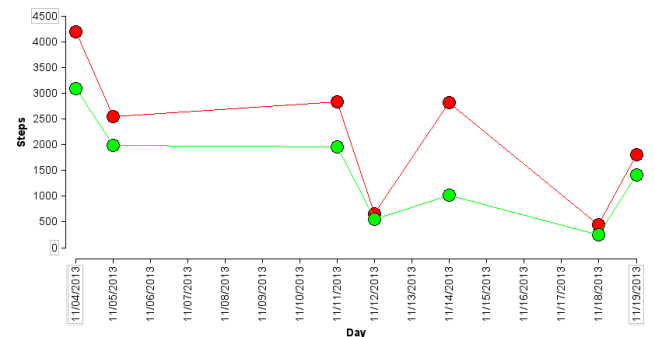

Figure 8. A single teacher's set of data from days with both Fitbit data (red/top) and iPhone app data (green/bottom).

Using data displays she had created in *TinkerPlots*, like the one shown in Figure 8, Emily began to interpret her data. First, she had a sense that the numbers being recorded were different but generally followed similar patterns. The range of values was more or less reasonable and could be explained given her awareness of what the typical teacher afternoon looked like:

> Emily: [Pointing to the data] **here it says 2000 that seems about correct for [the afternoon] because** she is sitting down at lunch, [then] we are sitting during the day during read aloud, also **my teacher has a lot of sitting and explaining** for a half an hour and we get to do the project for 10 minutes, so that is what we are doing most of the time.

Note that the awareness of standard afternoon activities, both from her own experience as a student and given her familiarity with what the teachers were often doing, played an important role. Having studied data from her own activities and those of her

classmates, Emily was able to calibrate her expectations for how many steps a student should get in a day. Her familiarity with the teachers' behavior, where their activities were similar to student activities and where they differed, allowed her to estimate a "correct" number of steps for a teacher's afternoon. This is one of the powerful uses of wearable tracking devices by children—it positions them as experts about what was being recorded. Most often, it involves the expertise of their own activities, but as schools operate with fairly standard routines for all people present, students also had some expertise about what their teachers did.

On Figure 8, a particular day stuck out as being especially concerning. The points on the first day for both devices seemed much higher than the points from every other day.

> Emily: [Pointing to the peak in Figure 8] **So right about here is the off day… she ran to a meeting**…that's only one time, she ran [extra] to her meeting and [extra] back.

In addition to noting the odd day in Figure 8, Emily also prepared a plot of all the data points from the two teachers and examined the distributions (**Error! Reference source not found.**). She noted that the Fitbit and the iPhone app had very different distributional patterns. The Fitbit data had a clear peak and tails on either side. The app had a concentration of points with lower values and some gaps in the distribution. Based on these, she also concluded that the kinds of numbers obtained from the app lacked consistency even though the teachers' daily routines were fairly consistent and should produce a normal pattern of variation. She attributed this to a number of things, including the necessity for the teachers to sometimes hold the iPhones in their boots or because they needed to use their phones during class time to look something up or respond to an urgent text message. Thus, her final conclusion was that the Fitbit was a superior tracking device.
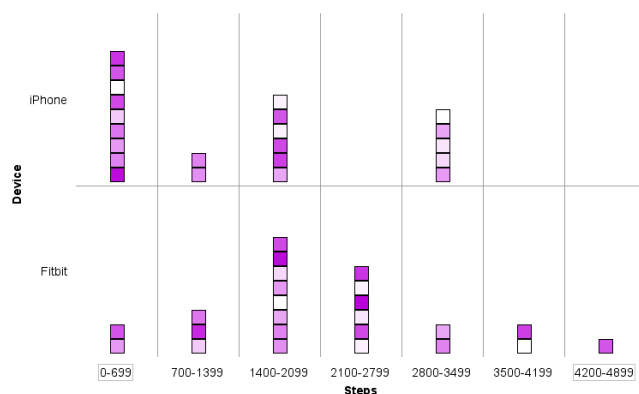
Figure 9. Distribution plots of all the step data. The data points from the app were organized in the top histogram on her data visualization window, while the data points from the Fitbit were organized in the bottom part of the window.

5.2　Things learned and demonstrated

When analyzing her teachers' data, Emily took into consideration what number of steps would be reasonable for a typical day. The numbers, which were in the thousands (of steps), could be explained based on the observations she had made on a daily basis of how much teachers moved around during the school day and knowing from her own and other students' personal tracking how much students moved around in the school day at different times. She had calibrated a sense of plausible values, similar to what we

have seen even when other students have very short periods of exposure to their self data [19].

She also demonstrated a useful understanding of distributional shape that went beyond describing shape in terms of canonical descriptors (such as mode and range) and used those to make inferences. When the underlying object being measured is fairly consistent (such as how much a teacher moves around in the afternoon at school), the pattern of variability should involve a peak and some tailing off. It should not have concentrations of data points in some areas and large gaps otherwise unless there was good reason to believe that the nature of the school day was inconsistent.

Finally, as in the first case with the soccer students, Emily realized that data collection was messy and could produce data points that may or may not be usable relative to a research question. She encountered this when she had missing or partial data from her teacher participants, and she came up with a reasonable solution for how to deal with this even though it affected the total amount of usable data in the end for her comparison.

5.3　Opportunistic Use of the School Day

Like the students in the previous case, Emily participated in class data analysis discussions during the Fitbit starters. Through these opportunities, Emily studied data from several weeks' worth of her own and others' daily activities and had learned what to expect of data from "routine" days as well as how to recognize the effects of non-standard activities and potentially flawed data. She was able to draw on these experiences when analyzing and interpreting the data in her investigation.

Emily's case is particularly interesting in that it illustrates how children can leverage aspects of their school days often overlooked by others. For most school children, the school day includes routine, close interactions many adults. Because of her relationships with some of her school's teachers, the movements of adults were a particularly salient portion of Emily's day. Through her volunteer work, she had steady access to and contact with Mrs. Thompson in her afternoons and had developed a sense of typical activities in her class, while also witnessing any deviations from the routine. Similarly, through interactions with her classroom student teacher, Emily came to understand her routines as well. She was able to opportunistically use this knowledge to eventually draw her own conclusion to a research project of her own design.

6.　CONSIDERING THE EFFECT OF THIS EXPERIENCE FOR THE LARGER CLASS OF STUDENTS

Emily, Geoff, Lauren, and Neill were not the only students to customize projects and analyze the school day. As a set of cases, they are appealing because we happened to get a good amount of data on their activities, and they were fairly articulate about what they had done when asked. In that year, students also looked at how far they had to walk to get back to the top of a nearby hill after sledding down using two different types of sleds; other students tried to determine if winning or losing students playing 3-on-3 basketball were demonstrably moving around the court more. Both of these were popular recess activities in the winter and in the spring, respectively.

Aside from some compelling cases, the question often arises in such explicitly education-oriented work about whether students made quantifiable gains in their understanding of targeted content.

In this case, we were interested in gains in statistical reasoning, an increasingly important set of competences that the Common Core State Standards are emphasizing in elementary and middle grades mathematics education. We were fortunate enough to be provided with a set of assessment items from a research team that has done extensive and quite respected work in the assessment of elementary student statistical reasoning from a "Learning Progressions" perspective [6]. Briefly, learning progressions are potential pathways informed by recent cognitive and educational research on the development that can be realized over extended time through well-designed and properly supported classroom activities [7]. They can serve as guides for curriculum design, teacher professional development, re-articulation of standards, and for student assessment. Ideally, instruction would have some vertical integration so that students would not only show incremental growth along a progression over a short period of time, but through multiple years of spiraling through the similar content and processes, they would demonstrate desired levels of scientific and mathematical sophistication targeted in the most recent standards documents [1]. These are knowingly far more ambitious targets than what has been done in math and science education in the past decades, but have been demonstrated as attainable through educational design research [21,25].

Using the same, psychometrically validated items that were developed, tested, and implemented in another study [21], we were able to quantifiably compare two fifth grade classes. One, which used Fitbit activity trackers and explored variability and custom projects first, was the source of the two cases presented in this paper. The other had not yet done the unit but would do it immediately after. They did have exposure to some elementary statistical content by virtue of what their teacher had opted to introduce to them at that point in the school year. Both were tested when the first class had finished the unit, so that we could compare the effects of this experience against classroom business as usual. The second class was tested a second time after they finished their instantiation of this experience. Unfortunately, because of timing and our ability to receive the testing items relative to when the first class could work with us, we do not have pre-test data for them. However, we have no reason to suspect, based on conversations with school administrators, that the students were grouped into different classes because they differed substantially in ability.

Assessments for each student were scored using multi-level rubrics designed to provide a fine-grained understanding of the performance thresholds achieved by each student across several knowledge and performance constructs. We then assigned numerical values to each threshold score, which we used to estimate performance gains for each classroom (see Table 1).

Table 1. Student performance on pre- and post-tests.

	Pre-test		Post-test	
Class	Mean	Std Dev	Mean	Std Dev
1			1.69*	1.37
2	1.03	1.24	1.62*	1.47

* statistically significant (p<0.05) compared to class 2 pre-test.

These results show that there was significant improvement in both classes. As learning progressions are a rather new and ambitious model for designers of educational experiences, these positive results are quite encouraging. They demonstrate that while students can come up with some novel ideas for individual projects situated within a school day, there are measurable aggregate changes of consequence. While more work remains to be done, these aggregate changes along with the cases we have presented suggest that the general approach of using wearable activity tracking technologies, focusing on familiar aspects of the school day, and stressing the collection and analysis of real data has promise and a possible home within the classroom.

7. REFLECTING ON OPPORTUNISM AND INTERACTION DESIGN FOR CHILDREN

Much of this paper has had an educational flavor and focused on kids learning about and doing statistics. Yet throughout this paper, there has also been a recurring theme related to designers and children both being opportunistic. Taken in its colloquial sense, being opportunistic can easily be seen as having a negative connotation. In this sense, an opportunist takes advantage of circumstances and others' hard work for their own, selfish advancement. Opportunistic innovation, then, becomes a product of serendipitous circumstance rather than of particularly creative individuals or groups.

Yet, there is little reason for us to see opportunism in design as an inherently negative thing. In interaction design research, some have already begun to re-appropriate the term in more positive ways. For example, "opportunistic design" has been proposed as a term describe the practices of hackers [2,9]. Hackers are observed as appropriating existing hardware and software components with which they are already familiar. They then modify and combine these components to create a new system that meets the hacker's functional requirements. Opportunism allows the hacker to leverage available resources and expertise in new and innovative ways without the expense and difficulty of generating a product with the same functionality from scratch. Why re-invent the wheel when we could cobble together a functional car with round objects that exist already?

We recognize there is an allure of creating radical new tools and experiences for children in radical new settings, and we are proponents of continued work in that direction. However, we contend much remains to be said about what can be done in familiar environments by appropriating tools that already exist. This opportunistic approach applies even in situations where an initial appraisal would suggest there are simply just an abundance of constraints that could limit innovation. This is often seen as the case for children's school days, where the schedules are rigid and routines are well established.

Yet we can design new experiences by seeing those constraints as resources and by considering how we can be opportunistic with respect to how those resources are used. For us, introducing wearable devices to children was key in this regard. While wearable devices have been quite popular in public media and also in HCI and ubiquitous computing circles, their use with children has been quite limited. (There are some noteworthy exceptions [24]). Through the cases presented in this paper, we have shown how wearable devices can be made useful to children and in the confines of a school day. Using wearable devices to collect data allowed us to leverage the constraints of the school day and expose them to the students' inspection. The students, in turn, used these opportunities for inspection to facilitate the development of questions relating to familiar activities.

The larger point we have sought to make through this paper is that while the school day can create a set of constraints on what and how we design experiences for children, there are also ways in which we can and should be opportunistic. We can find existing routines where specific data discussion activities can take place

(such as with the school's morning meeting). The repetitive routine of the school day also allows students to recognize what is standard in their experience and focus on where there might be meaningful variability (i.e., recess) subject to empirical testing and investigation. Considering half of their waking time on weekdays is spent in school, it might be wise for interaction designers to tap into the children's school day routines as a design resource.

And it is not just the routines of students that can be made into objects of inspection. Students know what others in the school building—including teachers—are doing as well. That can, as it did for Emily, be a driver and support for personal inquiry. In both of the cases we presented, the regularity of the school day also positioned the children as experts. The youth were able to make the kinds of determinations that scientists and serious data analysts must make such as what were the criteria for data to be kept in an analysis and what made data unusable relative to the questions that they are asking. The familiarity they had with what happens each day at school, we contend, played prominently in making that possible. In that respect, the students became opportunistic as well, and in our view is that was ultimately a good thing with respect to their learning.

And while wearable activity trackers were our tool of choice in this project and in this paper, we see little reason for this brand of school day opportunism to be restricted to wearable technologies. Imagine what could happen if the hallway foot traffic that naturally takes place between classes could be measured with pressure sensitive carpeting and used by a physics class to examine and measure forces exerted by (or against) people? Or what could happen if electricity meters were attached to standard school appliances such as projectors, desktop computers, fluorescent lights, and other frequently used school electronics as a vehicle to explore everyday energy consumption and conservation? These are just a few possibilities, but they, along with the cases of wearables that we presented above, hint at the possibility that there could be much to harness out of seemingly mundane activities and places. Approaching design for children with an eye toward opportunism may make those and other resources more visible to us and other designers in the future.

8. REFERENCES

[1] Achieve, Inc. *Next generation science standards.* Washington, D.C, 2013.

[2] Brandt, J., Guo, P.J., Lewenstein, J., Klemmer, S.R., and Dontcheva, M. Writing Code to Prototype, Ideate, and Discover. *IEEE Software 26*, 5 (2009), 18–24.

[3] Ching, C.C. and Foley, B.J., eds. *Constructing the Self in a Digital World.* Cambridge University Press, New York, 2014.

[4] Ching, C.C. and Schaefer, S. 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*. Routledge, 2015, 201–219.

[5] Ching, C.C., Wang, X.C., Shih, M.-L., and Kedem, Y. Digital Photography and Journals in a Kindergarten-First-Grade Classroom: Toward Meaningful Technology Integration in Early Childhood Education. *Early Education and Development 17*, 3 (2006), 347–371.

[6] Duncan, R.G. and Hmelo-Silver, C.E. Learning progressions: Aligning curriculum, instruction, and

assessment. *Journal of Research in Science Teaching 46*, 6 (2009), 606–609.

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

[8] Enyedy, N., Danish, J.A., Delacruz, G., and Kumar, M. Learning physics through play in an augmented reality environment. *International Journal of Computer-Supported Collaborative Learning 7*, 3 (2012), 347–378.

[9] Hartmann, B., Doorley, S., and Klemmer, S.R. Hacking, Mashing, Gluing: Understanding Opportunistic Design. *IEEE Pervasive Computing 7*, 3 (2008), 46–54.

[10] Kafai, Y.B., Fields, D.A., and Ito, M. *Connected Play: Tweens in a Virtual World.* The MIT Press, Cambridge, Massachusetts, 2013.

[11] Kafai, Y.B., Peppler, K.A., and Chapman, R.N. *The Computer Clubhouse: Constructionism and Creativity in Youth Communities. Technology, Education--Connections.* Teachers College Press, 2009.

[12] Kanter, D.E. Doing the project and learning the content: Designing project-based science curricula for meaningful understanding. *Science Education 94*, 3 (2010), 525–551.

[13] Konold, C., and Miller, C. (2005). *TinkerPlots Dynamic Data Exploration: Statistics software for middle school curricula.* Emeryville, CA: Key Curriculum Press.

[14] Konold, C. and Pollatsek, A. Data Analysis as the Search for Signals in Noisy Processes. *Journal for Research in Mathematics Education 33*, 4 (2002), 259–289.

[15] Lee, V.R. Students' Digital Photography Behaviors during a Multiday Environmental Science Field Trip and Their Recollections of Photographed Science Content. *Education Research International 2014*, (2014), 1–11.

[16] Lee, V.R. Looking at how technology is used with bodies over there to figure out what could be done with the technology and bodies right here. In V.R. Lee, ed., *Learning technologies and the body: Integration and implementation in formal and informal learning environments*. Routledge, New York, 2015, 167–184.

[17] Lee, V.R. and Briggs, M. Lessons Learned from an Initial Effort to Bring a Quantified Self "Meetup" Experience to a New Demographic. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM (2014), 707–710.

[18] Lee, V.R. and Drake, J. Quantified recess: Design of an activity for elementary students involving analysis of their own movement data. *Proceedings of the 12th International Conference on Interaction Design and Children*, ACM (2013), 273–276.

[19] Lee, V.R. and DuMont, M. An Exploration into How Physical Activity Data-Recording Devices Could be Used in Computer-Supported Data Investigations. *International Journal of Computers for Mathematical Learning 15*, 3 (2010), 167–189.

[20] Lee, V.R. and Thomas, J. Integrating physical activity data technologies into elementary school classrooms. *Educational*

Technology Research and Development 59, 6 (2011), 865–884.

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

[22] Lyons, L., Silva, B.L., Moher, T., Pazmino, P.J., and Slattery, B. Feel the Burn: Exploring Design Parameters for Effortful Interaction for Educational Games. *Proceedings of the 12th International Conference on Interaction Design and Children*, ACM (2013), 400–403.

[23] Moher, T., Ching, C.C., Schaefer, S., et al. Becoming Reflective: Designing for Reflection on Physical Performances. *ICLS 2014 Proceedings*, (2014), 1273–1282.

[24] Norooz, L. and Froehlich, J. Exploring Early Designs for Teaching Anatomy and Physiology to Children Using Wearable e-Textiles. *Proceedings of the 12th International Conference on Interaction Design and Children*, ACM (2013), 577–580.

[25] Petrosino, A.J., Lehrer, R., and Schauble, L. Structuring Error and Experimental Variation as Distribution in the Fourth Grade. *Mathematical Thinking and Learning 5*, (2003), 131–56.

[26] Rosebery, A.S., Warren, B., and Conant, F.R. Appropriating Scientific Discourse: Findings from Language Minority Classrooms. *The Journal of the Learning Sciences 2*, 1 (1992), 61–94.

[27] The Design-Based Research Collective. Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher 32*, 1 (2003), 5–8.

[28] Tinker, R. A history of probeware. 2000. http://concord.org/publications/articles-and-papers.

9. ACKNOWLEDGMENTS