EVALUATION OF TRANSPORT MODE USING WEARABLE SENSOR DATA FROM THOUSANDS OF STUDENTS

1 Erik Wilhelm (corresponding author)  
2 Singapore University of Technology and Design, EPD  
3 8 Somapah Rd., 487372, Singapore  
4 Telephone: +65 6499 4606 Email: erikwilhelm@sutd.edu.sg

5 Don MacKenzie  
6 University of Washington, Civil and Environmental Engineering  
7 201 More Hall, Box 352700  
8 Seattle, WA 98195-2700  
9 Telephone: 206-685-7198 Email: dwhm@uw.edu

10 Yuren Zhou  
11 Singapore University of Technology and Design, EPD  
12 8 Somapah Rd., 487372, Singapore  
13 Email: yuren_zhou@sutd.edu.sg

14 Lynette Cheah  
15 Singapore University of Technology and Design, ESD  
16 8 Somapah Rd., 487372, Singapore  
17 Telephone: +65 6499 4740 Email: lynette@sutd.edu.sg

18 Nils Ole Tippenhauer  
19 Singapore University of Technology and Design, ISTD  
20 8 Somapah Rd., 487372, Singapore  
21 Telephone: +65 6499 4869 Email: nils_tippenhauer@sutd.edu.sg

22 5207 words + 8 figures + 1 tables = 7457

23 August 1, 2016
The main question we address in this paper is: “What role does travel distance, mean speed, and walking distance play in Singaporean student travel?”. This is important for dense urban cities such as Singapore where congestion problems are growing. We use data that was gathered by 43,000 students and offers unique, unbiased, door-to-door commuting trips. Each student was wearing a sensor that tracks location and physical parameters. At the beginning of this analysis, our hypotheses were that (a) private transportation modes are substantially faster, more convenient, and more comfortable than public modes; (b) people living close to public transportation infrastructure will use it preferentially; and (c) students choose routes that minimize walking distance even if it increases their overall travel distance. These hypotheses were tested by examining descriptive statistics about the data set which we present together with a novel approach to comparing observed travel behavior with Google Directions suggestions. We find that contrary to our first two hypotheses, students traveling in private cars are less than 10km/hr faster for their door-to-door commute than their peers who use public transport. Students living close to metro stations do not necessarily choose to travel by train despite a very comprehensive and convenient network. We find and present some evidence to support our intuition that Singaporean students will make choices that minimize the amount of walking which they must do in their morning commute.
INTRODUCTION

The challenge of obtaining accurate door-to-door transportation pattern data is well-known in the transportation research community. Surveys may be comprehensive, but are subject to substantial bias and accuracy issues (1). Traffic loop and transit gate counting technologies may be accurate, and attractive to research at a macro scale, but offer neither comprehensive coverage nor the ability to capture full trips (2). There is a strong need, driven by growing urbanism and associated traffic congestion, for systems which can capture detailed and complete commuting trip information accurately and comprehensively from a broad swath of a city’s populace in order to design effective policy. Smartphone-based activity trackers have been used successfully in the past (3), but that approach always faces adoption and recruitment challenges. In that context, our work in sensor platform development which enabled the analysis in this paper is able to fill an important gap in transportation data, and our initial analysis answers some important pressing transportation policy questions relating to first/last mile commuting.

National Science Experiment

The National Science Experiment (NSE) is an initiative in Singapore with the aim to encourage school students to engage in engineering topics such as data analytics, computer science, physics and mechanics. As part of the NSE, 50,000 wearable sensors (called SENSg) were designed, prototyped, and built together with a major electronics company. 128 schools participated in the 2015 experiments for one week, during which students wore the SENSg devices on lanyards during normal daily activities. The SENSg then measures environmental data such as temperature, ambient light intensity, humidity, etc. as well as location at regular intervals. Data is stored locally on the SENSg and uploaded to a central cloud infrastructure when in range of a suitable Wifi network (e.g. the school’s network). Students can access their SENSg’s data via a web application that provides data point visualization on a map, interactive games and analytics, and allows them to download the sensor’s data.

In the context of this work, the most important features of the NSE are two-fold: each measurement can be localized with an accuracy of around 20 m by leveraging WiFi access points observed during the measurement (if at least 2 access points were observed). In addition, the data measured by the SENSg nodes allows us to estimate the current transportation mode of the user (4).

Background

Examining transportation mode choice is a well established active research area and the decision process is generally considered to be complex and multivariate (5). Travel time is an important predictor of mode choice and appears in most mode choice models. Comprehensive reviews and meta-analyses of the research literature show that the value of travel time is not constant: it is generally different for travel in different vehicle types, for walking, and for waiting time (6, 7). The NSE data can offer opportunities to validate and refine such models based on a rich, real-world data sets. This paper offers a preliminary look at mode choices inferred from the NSE data and links them to trip distance, proximity to transit services, and related data.

Examining preferences for walking and walkability in cities also has valuable implications for transportation system planning and urban design. A previous study on this topic found that pedestrians in Singapore tend to walk less than pedestrians in other cities (8). Our work helps to clarify some of these decisions in the Singaporean context and to add a powerful method for future researchers interested in similar questions.
Contributions
Our contributions to the state of the art in understanding the commuting mode decisions of students in dense urban environments are:

1. A novel approach to collecting student travel data on a very large scale using custom-designed wearable sensors,

2. An analysis of the data set to address specific questions about student travel behavior,

3. A unique approach to investigate mode decisions by comparing Google route suggestions with observed travel behavior,

4. Conclusions regarding the mode choice in relation to mean commuting speed, proximity to transportation services, and tendency to choose modes with based on amount of walking required.

METHODOLOGY
Our main research questions focus on the factors influencing travel mode choice for students with particular emphasis on the first/last mile characteristics of their trips. We define the first/last mile “zone distance” as 500m from the home or school to the nearest bus stop, and 2000m to the nearest train station. While in Singapore the train system is referred to as an “Mass Rail Transit” (MRT), we will refer to it as “metro” or “train”. This section describes the pre-processing of the data, how it was prepared for presentation, and the sensitivity of the results to our methods. The 2015 experiment was performed by 43,140 students in Singapore. The ages of these students were not explicitly considered in the analysis, but the participants were from 54.5% primary (age 7-12 years old), 40.5% secondary (13-16 years old), 5% Junior College (17-18 years old) levels. After processing the time-series data to identify home-school location, travel mode, and filtering to ensure that sufficient data is present, a set of 2852 students with excellent data remain to be analyzed. Morning commutes from home to school are the only travel considered due to the variability introduced by students participating in a range of after-school activities. Except for the mean velocity study, the cleaned data used in this paper contains 510 walking, 1527 bus, 285 train (metro), and 530 car commuters. In order to focus on students exhibiting routine behavior, students were only included if they took repeated trips of the same mode, and for 21% 1 trip, 41% 2 trips, 29% 3 trips, 8% 4 trips, and <1% 5 trips were correctly identified on subsequent days. For components of the analysis where a complete set of features is required, a further subset of 1335 students are studied, and this reduction is caused by a lack of comprehensive ground truth data from Google directions queries. In order to determine the nearest bus stop or metro station a zone distance from home or school, an R-tree model of Singapore’s transportation network was trained based on publicly available station information.
In this work we do not control for or consider socio-economic factors relating to car ownership, the dangers of walking and cycling in Singapore, the influence of age or gender on the mode or route decisions, as this data was not collected on a sensor-by-sensor basis in 2015 due to privacy considerations. To evaluate how students choose their commuting transport mode, we compare how they actually commute given the transport options they have. We divided this procedure into two steps:

1. Process raw data to get attributes of the real trip, including locations of origin and destination, total travel time, total travel distance, total walking distance, and our estimate of student transport mode (walking, bus, train, or car),

2. Query Google Directions API to obtain the route options for different public transport modes.

**Step 1: Trip and Mode Identification**

Using the SENSg device, anonymous data is collected from students with a sampling frequency of 1/13 Hz, including timestamps, geographical mobility data (coordinates), inertial measurement units (IMU) data (accelerometer, magnetometer, gyroscope), environmental data (light, temperature, pressure). From the timestamps and geographical coordinates, approximate speed can be calculated for each sample. By applying a velocity-based Place of Interests (POI) detection algorithm, important locations such as home and school can be identified. Details on this algorithm were published here (10). After identifying the home and school location, the morning trip from home to school can be easily sliced from the data and aggregate attributes of the morning trip (such as travel distance and duration) calculated.

We developed a unique transport mode identification algorithm to extract the true mode of the given trip, both at a point-level and trip-level. Point-level mode refers to the transport mode of a particular data sample and trip-level mode refers to the aggregated mode of a given trip. The algorithm is based on a Decision Tree model, leveraging features calculated from IMU data and geographical data. The details of this algorithm can be found in our previous publication (4). Validation of this algorithm is presented in the [Automatic Travel Mode Labeling] section.

**Step 2: Transport Mode And Route Queries**

To obtain all possible transport choices for the students, the Google Directions Application Programming Interface (API) (11) is used. This API has been applied by many previous research groups investigating a variety of topics. For example, it was adopted as the basis for a more advanced map service (12), it supported implementation of customized navigation system (13), it is also proved to have advantages in estimating travel time given pair of end points compared with traditional methods (14). Most relevant to our work is the previous comparison of API suggestions with the real routes taken by drivers (15).

The Google Directions API requires four inputs: trip origin, destination, transport mode and departure time. The origin for students’ morning trips are their home location and the destination is their school location identified in the first step. Because the departure time used must be a future time, the “hour:minute:second” uses same values as the real departure time, while the “date” is carefully chosen such that it is in the future and also has the same day of week as the real departure time. We then query Walking, Driving, Transit modes individually to get all possible
Google routes. Usually these result sets contain more than one route, and we therefore record each possible route for the given mode.

Inside the results returned by the API there are discrete data points as well as aggregated attributes of each possible route, including geographical coordinate list of points along the route, total travel distance, total travel duration. We process that data to obtain attributes such as total waiting time, total walking distance of each possible route.

Automatic Travel Mode Labeling

To make sure the identification of the transport mode of each trip is correct, we developed a novel automatic ground truth labeling (henceforth referred to as auto-labeling) system based on a comparison of observed travel and Google Directions results. The system can evaluate both point-level and trip-level modes. The modes identified by algorithm described by our earlier publication (4) are then compared with the modes labeled by this auto-labeling system, and only when these two types of trip-level modes are consistent, the trip is added into the data used for the analysis in this paper. In other words, we try very hard to only include modes for analysis in this work when they are considered to be accurate.

In the auto-labeling system, routes of all possible modes returned by the Google Directions API are used as references to label the true modes for students’ trips automatically. The basic idea behind this label validation method is that when people travel with different modes, the attributes of their real route (geographical shape, distance, duration, etc.) are significantly different. So if there is one route returned by the API which is similar to the real trip, then the real trip is very likely to have been made using the same transport mode as the API route does.

Figure 1 illustrates how this idea works. In the map, cyan and afternoon dots stand for morning and afternoon trip of a student respectively, red, blue, green and yellow lines stand for Google’s bus, metro, driving and walking routes. It is obvious that the morning trip follows the bus route while afternoon trip follows the metro route. Note that in the results we present we only ever consider the morning trip to avoid the variability present in evening travel. Based on this idea, we label the true modes of trips after data collection automatically, saving manpower and avoid recording errors. The auto-labeling algorithm follows these steps:

1. Compare each Google route with the real track and calculate similarity features which guarantee that the two are similar in all aspects (mainly geographical shape, distance and duration),
2. Pick up the Google route which is most similar to the real track according to the similarity features,
3. Check whether the similarity features are within certain thresholds which are trained from historical data,
4. If step 3. is passed, then the selected Google route is a trustworthy reference. Its mode is then taken as the true mode for the real trip. All data samples are correspondingly labeled.

By comparing the point-level modes labeled by this automatic method with ground truth modes collected manually, we compose the performance matrix shown in Table 1.
FIGURE 1: Example of judging the true modes of a trip offline with reference to routes of different modes from Google Map: cyan and orange dots represent morning and afternoon trip of a student respectively, red, blue, green and yellow lines represent Google’s bus, metro, driving and walking routes. It can be observed that the morning trip follows the bus route, while the afternoon trip follows the SMRT route.

The table is a good measure of a binary classification’s accuracy and is calculated as \( F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \). A total accuracy of 94% is observed for the over 6000 test points used. From the performance table and an understanding of the method, there are mainly four shortcomings of this auto-labeling system that we want to highlight:

1. If the real trip track doesn’t match with any Google route, the trip cannot be labeled. Intuitively, this happens most often for walking and driving trips. Since the focus of our present work is mainly on students’ choices on public transport modes, we tolerate this shortcoming but acknowledge it must be controlled for in further analysis,

2. Our system cannot differentiate between walking and stationary states, and therefore cannot be used as a feature in mode identification, but rather only as a validation tool,

3. Distance errors are cumulative, and our localization is 10-20m accurate at each point which can lead to large cumulative error. Thankfully these errors are strongly correlated and in-house ground-truth tests have shown that the travel distances are ultimately not subject to excessive error,

4. Small errors in sample-by-sample mode identification during mode transitions are common i.e. walking to get on a bus. The evaluation of transport mode choices we present
TABLE 1: Performance Matrix for our Auto-labeling System

<table>
<thead>
<tr>
<th>Mode</th>
<th>Precision</th>
<th>Recall</th>
<th>F₁ score</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking/Stationary</td>
<td>0.84</td>
<td>0.73</td>
<td>0.78</td>
<td>895</td>
</tr>
<tr>
<td>Metro</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>435</td>
</tr>
<tr>
<td>Bus</td>
<td>0.94</td>
<td>0.97</td>
<td>0.95</td>
<td>3617</td>
</tr>
<tr>
<td>Car</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>1752</td>
</tr>
</tbody>
</table>

Summary: **Accuracy:** 0.94  6699

here does not require such detailed classification, and hence this shortcoming is also non-critical.

We endeavor to only include students for analysis when we feel we have accurately identified their commuting modes, but acknowledge that no system is perfect. To verify the impact of potential mis-classification on the results, we test the effects of randomly selecting and swapping students from one class to another in the data set on mean velocity results. The procedure is relatively straightforward after we swap the velocities from one class to another we re-run the statistical tests of interest and examine if the results change. This allows us to explore how robust the differences in mean velocity are to failures in classification.

The mode assignment from point-level to trip-level is finally performed by examining the overall trip, and finding which mode in which the student traveled the greatest distance. This is then assigned to the student as its mode. Students are only included if their “main mode” is consistent through the days of the experiment for which we have data in order to study students with habitual and constant travel choices. Variation between mode choices through the week is an interesting topic for further study.

Processing API Queries
In order to compare Google’s API response to what we observe, we calculate the following parameters:

- trip duration $G_{dur}^{tot}$ (seconds)
- Google trip distance $G_{dist}^{tot}$ (kilometers)
- waiting time (seconds)
- number of transfers
- walking distance $G_{walk}^{dist}$ (kilometers)

The Google API typically returns 1-4 options for transit routes between each origin/destination pair which we send. From these, we calculate min, mean, and maximum values which we utilize in the analysis in the Results section. The approach we use to calculate the difference between measured ($M_{dist}^{tot}$) and Google total trip distance for $N$ Google API responses is shown in Equation 1, and the difference for walking distance we calculate is shown in Equation 2. To parameterize travel...
choice variation between available options and chosen routes, the minimum differences between
Google API suggestions and measured quantities are calculated using simple differences, and are
annotated as \(D_{MIN\text{dist}}^{\text{tot}}\) and \(D_{MIN\text{dur}}^{\text{tot}}\) for the minimum distance and duration returned by Google
respectively, assuming one trip was the actual trip taken.

\[
D_{\text{dist}}^{\text{tot}} = \frac{1}{N-1} \sum_{i=1}^{N-1} (G_{\text{dist}}^{\text{tot}}) - M_{\text{dist}}^{\text{tot}} \tag{1}
\]

\[
D_{\text{walk dist}}^{\text{tot}} = \frac{1}{N-1} \sum_{i=1}^{N-1} (G_{\text{walk dist}}^{\text{tot}}) - M_{\text{walk dist}}^{\text{tot}} \tag{2}
\]

5 RESULTS
The results shown in this section draw only from the subset of the National Science Experiment
data described in the methods section for which the authors are confident that the mode identi-
fication has been made as accurately as possible. Throughout this section, sources of error and
implications of mis-classification to the analysis are highlighted and discussed. In the first section
we discuss how the mode chosen by students impacts their mean travel speed. In the second sec-
tion, we present the results of an analysis on the impact of how far a student lives from school or
transportation system stations on their mode choice. In the final section, we examine the differ-
ences between route options offered by Google’s Directions API and the actual distance traveled
to test the hypothesis that students prioritize limiting walking distance over total travel distance.

15 Relationship between Mode and Mean Speed
The increase in mean travel speed moving from walking to personal vehicles is shown in the box-
whisker plot in Figure 2, which makes it clear that, as expected, vehicular transportation is faster
than walking. Interestingly, the measured mean speed of public transportation is only 6 and 9
km/hr slower than cars for train and bus modes respectively, even though the public mode speeds
include waiting and transfer times. The walking, bus, train and car mean speeds are 6.9, 13.0,
16.4, and 22.2 km/hr respectively, and the distributions of walking speeds can be seen in Figure
3. Compared with a normal curve with the distribution’s mean and standard deviation the mean
speeds appear to be normally distributed with the exception of walking. In order to determine the
significance of these differences, a Tukey-HSD test was performed on the log of the mean speeds
which brings the variance to within 0.2 log(km/hr). The results of the Tukey test indicate that
there are significant differences in the mean speeds of all of the modes, which is confirmed by
finding that the standard error of the mean speeds are less than 10% of the differences between
the means. These differences are stable even when 75% of the total population are mis-classified
using the methods outlined in the Automatic Travel Mode Labeling section. The authors note
that Figure 2 suggests that traveling by private vehicle in Singapore during morning commutes is
not substantially faster than using public transportation, with bus trips being the slowest and most
variable. Our measured walking speeds agree with previously published results which suggest that
walking speeds are between 4-4.5 km/hr (8). School start times are very consistent, and congestion
during school starting hours has been well documented and road pricing adjustments have been
FIGURE 2: The box/whisker plot shows how the median trips taken by car are only 6 to 9 km/hr faster than those taken by bus and train respectively. Measured walking speed aligns well with previously studied and published values.

applied as counter-measures \((16, 17)\). Students who have access to a car are likely using this mode due more to comfort and convenience factors rather than the desire to minimize travel time.

3 Travel Distance and Transport Mode

In this section we explore mode choice as a function of distance from home to school. Public policy gives primary school students preferential enrolment to schools within a shorter distance from their homes, and to examine the implications of this policy we averaged the median travel distance for all nodes at a particular school and plotted them for the 128 participating schools in different levels in Figure 4. The median home-school distance for all nodes at a given school is lower for primary students than those attending secondary schools or junior colleges (JC’s); consequently, primary students tend to walk more than their older counterparts. In this study we did not investigate the impact of age on mode choice in detail and leave that as future work. Subsequently collected data from 2016 and 2017 will contain more comprehensive demographic data.

The first indication which suggests students attempt to minimize walking distance (and hence increase comfort due to reduced perspiration) at the expense of other factors such as travel time, reliability, etc comes from the features of the shape of the distribution in Figure 5. We see from this figure that half of the students who choose to take a bus live within 800m of a train station. Only half of the students live within 400m of a train station will choose this mode, which suggests that students are unwilling to walk additional distance to take the train. This is in accordance to the evidence collected by \((8)\) that pedestrians typically walk 300m and strongly prefer shortest walking routes above other factors. This evidence, while compelling, is not conclusive since it does not consider the attractiveness of the options available to the students, e.g. the bus stop near
FIGURE 3: The distribution of mean speed for students using various modes of travel are relatively normal with the exception of walking modes, and have standard deviations which are sufficiently comparable to apply a Tukey-HSD test.

FIGURE 4: Primary students travel the shortest distances and secondary and JC students travel the longest distances to reach their schools as shown here for the morning travel for each of the students traveling to the 128 schools in this study.
The mode share as a function of distance from home to school, as well as distances to the nearest bus stops and train stations are plotted in Figure 6. This once again shows that despite proximity to train stations, many students opt to take a bus, strengthening our previous argument. There are other interesting observations which can be made about this figure. It also shows that more than half of the students living within a 1km of their schools will walk. The authors note that the likelihood of students who live more than a few km from school choosing walking as their mode is small, and explain the small fraction seen at greater distances as a) mis-classification of mode, and b) students using bicycles, e-bikes, scooters, or other forms of personal mobility which the sensors were not trained to identify. As expected, train mode share increases as home-school distance increases up to 8km, but then loses share to bus up to 11km (the furthest distance in the set). This unexpected trend does not have a clear explanation, but could be related to how the Singaporean metro network is laid out, or how the current mode is classified as 'main mode'. This trend should be studied in more detail, ideally by breaking trips down into sections by mode. Additionally, some large fraction of students will travel to school in private school shuttle buses, and these trips would be expected to be taken independent of home proximity to train stations.

Figure 6 shows that car mode share is relatively insensitive to any of the distance metrics, which was also observed in previous trials, and explained by the fact that for morning trips, many parents who are car-owners will drop their children off at school before work. As expected, the students living very near a train station are much more likely to choose this mode. Interestingly, the distance from the nearest bus stop does not appear to have a strong effect on mode share; in fact, counter-intuitively, bus mode share increases as distance between home to the nearest bus stop increases, indicating that there are more complex geo-spatial effects here which we have not...
FIGURE 6: The figure shows that the farther someone lives from school, the more likely they are to take a train up to a distance of 8km, then bus regains share; that as the distance from home to the nearest bus stop increases there is an increase in the car mode-share at the expense of train and bus up to 250m, then bus regains share; and that as distance between school and the nearest metro increases there is a sharp decrease in the likelihood that a student will travel via train. The histogram shows that trips vary 1-4km from crows-flight distance between home and school. Negative numbers are mis-identification of home/school location, and a few percent of trips are affected.
yet analyzed. The shape of the comparison between crows-flight and actual distance traveled indicates that the transportation system demands some additional travel, between 1 and 4km typically. Negative differences are indicative of numerical errors in the geo-spatial data, and are thankfully a small fraction of the total set.

**Students’ Proclivity to Walk**

In this section, we present to answer the main question posed “do students act to minimize walking over other factors in their mode choice decision”? Using the methods described in section [Automatic Travel Mode Labeling](#), the difference between the mean of up to four Google Directions recommended routes are calculated using the measured distance traveled by foot and using public transportation. It is assumed that Google directions minimize travel time. The results presented in this section completely exclude walking and car trips, since these are not relevant to answering the primary question. Figure 7 shows that well over half of the students will walk a shorter distance than Google recommends for their morning trip between home and school. A substantial fraction will also travel a shorter distance overall, which indicates that Google may not capture all options for students traveling between these points. As shown in Figure 8, this trend also holds (although weakens) when we considered the minimum API distance versus what was measured. These results must be analyzed with the understanding that our sensor localization is not perfect, and that errors are additive in the measurement of distance.

A heat map of the differences between the mean of Google’s suggestions for walking and total trip distances versus what was actually traveled is shown in Figure 7. The quadrants of this heat map are labeled depending on the fraction of students present in each. 63% of the students walk less than Google suggests to varying degrees, and the average walks 315m less, and travels a total of 235m less. The heat map clearly indicates that the majority of students tend to travel closer to the total distance suggested by Google, but will walk less than the suggestion. The authors suspect that the reasons for this difference are a) the students access short-cuts through estates etc which are not documented by Google maps b) localization error may not capture all of the trip c) students are choosing to take short-hop buses which may not be in Google’s set of suggestions instead of walking.

To investigate this phenomenon further, the data was split according to mode choice and plotted as a histogram in Figure 8 for both the mean and the minimum Google travel distance result. This figure illustrates that the difference between the mean walking route distance suggestion and actual travel differs between students traveling by train and by bus. Students walked 440m less than Google predicted if they took the bus, and 90m more than Google predicted on average if they took the train. This strengthens the hypothesis that short-hop buses are being taken, as 440m is a substantial distance and above what could be expected to be explained by estate short-cuts or localization error. Further evidence that students are choosing routes which are light on walking comes from examining the difference between minimum walking Google route suggestion and what was actually traveled in Figure 8. This shows that in fact bus trips are still slightly negative on average, but train trips walk almost at Google’s suggestion.

Distance and time are strongly correlated when comparing Google Directions results which assume mean velocities for various legs, but it is nevertheless interesting see that students also tend to arrive faster than the Google recommendation, an average of 1 minute for train trips and 3 minutes for bus trips. The authors suspect that Google’s trip duration estimates are intentionally conservative. In general, the observed travel time estimates fit Google’s estimates closely.
FIGURE 7: Students taking public transit walk a mean of 314m less than Google’s route recommendation, and travel a total distance with a mean of 235m less that Google’s recommendation. This indicates that students are likely prioritizing factors beyond travel distance in their decisions. The ratios shown in the figure indicate that 63% of students walk a shorter distance than Google predicted; 29% walked a shorter distance but traveled longer.

FIGURE 8: Students who chose to travel by bus walked an average of 440m less than the mean of the walking distances in routes suggested Google; those who traveled by train walked an average of 90m more. This 530 m difference is statistically significant. Very interestingly, walking for train trips were 550m longer in walking distance than the minimum Google recommendation, whereas bus trips were 50m slightly shorter in walking distance than the minimum walking Google recommendation. On average, students are not picking the minimum walking distance trips for travel by train, which implies that their utility includes the reliability and speed of train travel.
CONCLUSIONS

In this paper we present a novel personal sensing system which we designed in the framework of the Singapore National Science Experiment project, show a series of descriptive statistics derived using a unique approach to obtain trip options from Google Directions API, and reach the following conclusions about student decisions in the Singaporean transportation system:

1. The mean door-to-door commuting velocity of students using public and private modes of transportation are separated by less than 10km/hr in aggregate, highlighting that the choice to travel by car in Singapore is likely a factor of the availability of this mode, as well as the comfort, convenience, and other features.

2. Proximity to transportation services alone is not sufficient to explain the share of different mode; in particular proximity to train stations does not predicate a tendency to travel by train, although train mode share increases together with trip distance.

3. Students tend to choose trips which minimize walking distance at the expense of greater overall travel distance.

Future work should include an analysis of socio-demographic data such as age, gender, income, and car ownership which were only available to the authors quite recently.

ACKNOWLEDGMENTS

This work was supported by the Singapore National Research Foundation, Ministry of Education, and the SUTD-MIT International Design Centre. On-time production of the sensors was only possible due to strong support from Delta Electronics DRC, IABG, and Taoyuan Factory 2. Logistics for the distribution of thousands of sensors was facilitated by the Singapore Science Center, and the infrastructure to securely store the data was built and maintained by the A*STAR Institute for High Performance Computing. The authors also thank Francisco C Pereira for the excellent discussion on developing the statistical components of this work.
REFERENCES


[9] Land Transport Authority (LTA), *Land Transport Authority Data Mall*. [https://www.mytransport.sg/content/mytransport/home/dataMall.html](https://www.mytransport.sg/content/mytransport/home/dataMall.html), 2016, last accessed July 1 2016.


