The Tesserae Project: Large-Scale, Longitudinal, *In Situ*, Multimodal Sensing of Information Workers

Stephen M. Mattingly[1*], Julie M. Gregg[2*], Pino Audia[3], Ayse Elvan Bayraktaroglu[4], Andrew T. Campbell[3], Nitesh V. Chawla[1], Vedant Das Swain[5], Munmun De Choudhury[5], Sidney K. D'Mello[2], Anind K. Dey[6], Ge Gao[4], Krithika Jagannath[4], Kaifeng Jiang[7], Suwen Lin[1], Qiang Liu[8], Gloria Mark[4], Gonzalo J. Martinez[1], Kizito Masaba[3], Shayan Mirjafari[9], Edward Moskal[1], Raghu Mulukutla[9], Kari Nies[4], Manikanta D. Reddy[5], Pablo Robles-Granda[1], Koustuv Saha[5], Anusha Sirigiri[3], Aaron Striegel[1]

[1] University of Notre Dame, [2] University of Colorado, Boulder, [3] Dartmouth College, [4] University of California at Irvine, [5] Georgia Institute of Technology, [6] University of Washington, [7] Ohio State University, [8] University of Texas at Austin, [9] Carnegie Mellon University

[*] *First and second authors. Other authors are listed in alphabetical order; Contact Author: smattin1@nd.edu*

ABSTRACT

The Tesserae project investigates how a suite of sensors can measure workplace performance (e.g., organizational citizenship behavior), psychological traits (e.g., personality, affect), and physical characteristics (e.g., sleep, activity) over one year. We enrolled 757 information workers across the U.S. and measure heart rate, physical activity, sleep, social context, and other aspects through smartwatches, a phone agent, beacons, and social media. We report challenges that we faced with enrollment, privacy, and incentive structures while setting up such a long-term multimodal large-scale sensor study. We

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3299041

discuss the tradeoffs of remote versus in-person enrollment, and showed that directly paid, in-person enrolled participants are more compliant overall compared to remotely-enrolled participants. We find that providing detailed information regarding privacy concerns up-front is highly beneficial. We believe that our experiences can benefit other large sensor projects as this field grows.

KEYWORDS

Sensors; social media; smartwatches; phone agent; stress; privacy

ACM Reference Format:

Stephen M. Mattingly et al.. 2019. The Tesserae Project: Large-Scale, Longitudinal, *In Situ*, Multimodal Sensing of Information Workers. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3290607.3299041

INTRODUCTION

Individual use of personal informatics systems like smartwatches has been growing in popularity. At the same time, researchers have been discovering the potential of wearable devices and other sensors for use as research tools to understand individuals in their natural environments at-scale. Some early efforts include projects such as NetHealth [4], StudentLife [7], CampusLife [6], and WorkSense [2] that share several key characteristics. First, they have used multiple sensors to measure different individual attributes such as mood, movement, and technology use. Second, they have time-synchronized sensor data to study relationships, such as how stress covaries with events.

The Tesserae project builds on these studies to investigate how a suite of publicly available commercial sensors can be leveraged to measure job performance (e.g., organizational citizenship behavior), psychological traits (e.g., personality, affect), and physical characteristics (e.g., sleep, activity). Traditionally, measures of these variables use self-reports which have known biases (e.g. [3]). These measures are generally taken only once or at a few points in time using surveys or experience sampling methods. The methodological advantages of using unobtrusive sensors are that the sensors can provide continuous, longitudinal measurements, in situ, and at scale.

In this paper, we report on the Tesserae Project, a large-scale yearlong study testing the feasibility of using sensors across a range of participants and contexts. We have enrolled 757 participants located across the U.S. who conduct information work (e.g., consultants, administrators, engineers, and managers). We report our challenges with enrollment, insuring privacy, and designing incentive structures to motivate long-term sensor use in varied contexts. Because Tesserae is unprecedented in scope, we provide reflections on our experience along with lessons learned to guide future researchers in similar endeavors.

METHOD: DESCRIPTION OF DATA COLLECTION

Participants

757 participants were recruited across the U.S. from a variety of university and corporate partners. Recruitment was done through workplace emails, messaging boards, and newspaper advertisements, with a Google Form link for interested participants to sign up. After screening participants, we emailed a detailed consent form (see Participant Privacy). This project was approved by the University of Notre Dame IRB and through reliance agreements at participating research universities.

Ground Truth Measures

Enrollment consisted of participants completing a set of initial questionnaires that served as ground truth measures for our sensors. These questionnaires assessed job performance, intelligence, personality, mood, anxiety, health measures, exercise, sleep, and stress. The initial ground truth battery (IGTB) took approximately 45-60 minutes to complete. For 56 days following enrollment, participants received daily text messages with links to abbreviated versions of the initial ground truth survey measures. For instance, the daily survey used a single question to assess sleep whereas participants completed the 20-item Pittsburgh Sleep Quality Index [1] as part of the initial battery. Each daily survey took about 3 minutes to complete. This combination of initial detailed measures and daily surveys provided measures of baseline and daily fluctuations in workplace performance, health, and well-being that serve as a ground truth to try to predict with our sensor suite.

Sensor- Wearable

Participants were provided with a Garmin Vivosmart 3 waterproof wristwatch with 4-5 days of battery life. The watch captured physical activity, Heart Rate (HR), sleep, calories, and floors climbed. Additionally, streaming data via Bluetooth provided beat-to-beat intervals (BBI), an HR measure useful for calculating HR variability, a physiological correlate of stress.

Sensor- Phone Agent

Participants installed a researcher-created app [7] on their smartphones, which tracked screen lock/unlocks, data usage, charging, location and other phone states. The phone agent also records beacon sightings (see Sensor- Bluetooth Beacons) and data streamed from the wearable. Battery drain was minimized, with iOS and Android apps consuming no more than 7% and 3% of the total battery life. Currently, 434 iOS users and 294 Android users are enrolled.

Sensor- Bluetooth Beacons

Participants were asked to place Bluetooth beacons in their offices and homes. using Gimbal Bluetooth Beacon Series 10 and 21 beacons (small transmitters). The beacons transmit signals captured by the phone agent. The sightings are timestamped and contain the beacon ID, signal strength, and current temperature. The Series 10 beacons are larger static beacons meant to be placed in the home and at work, with an 18-month battery life, allowing participants to place them and forget the beacons are there. The Series 21 beacons are key fobs slightly larger than a quarter and meant to be carried in a wallet, purse, or laptop bag and/or on a keychain. Beacon sightings can be combined to generate features (e.g. daily commute = duration between last home beacon sighting and first work beacon sighting on a work day). Beacons can also be used to study work interactions among study participants.

Social Media

Participants provided data from their Facebook and LinkedIn accounts. They could optionally provide Instagram and Twitter. For a detailed description of data collection and security measures, see [5].

FINDINGS: WHAT WE LEARNED

Participant Incentives

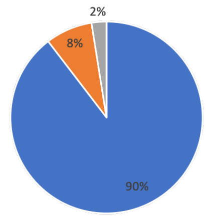
We faced a dilemma of how to incentivize participants because of the volume of data that we needed to collect, the potentially sensitive nature of the data, the longitudinal span of the study, and our limited resources. Ideally, we wanted participants to be intrinsically motivated to participate, i.e., to join the study because of their interest in its scientific value and also to see their own data patterns. However, we also recognized that we needed to extrinsically motivate participants, to compensate them for their time and efforts.

Our design decision was to use incentives that would be compliance-based. We considered participants to be compliant if they had data 80% or more of the time from their phone agent, their wearable, and completion of the daily surveys (on average across all three). We granted compliance amnesty to participants during active troubleshooting periods. Participants used a specially-designed portal to monitor their overall and individual stream compliance and could also report bugs and other issues via the portal. Based on corporate partner requirements, we designed two remuneration structures: direct payments and a lottery. For direct payments, compliant participants could receive up to $750. The payment schedule encouraged participant retention, with $25 for IGTB completion, $25 at the end of the first week, $150 at three months, $200 at six months, and $350 for completing the study. For each pay period, noncompliant participants received no payment. For the lottery, participants received a ticket per day for each stream (wearable, phone agent, daily survey) with 80% or more compliance. Thus, more compliant participants had higher odds of winning. Further, the number of

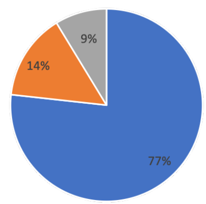
$250 drawings each week was proportional to the size of the participant pool. Lotteries were held every week. It is important to note that direct payment participants need all three streams to be recording data in order to meet the 80% benchmark, while lottery participants could still win with only a single stream of data for one lottery period. We observed that all-or-nothing payments led to more correspondence with participants via our compliance portal bug-reporting system or email and participants were more motivated to correct specific compliance problems in order to receive payments. Of course, this could also be attributed to differences across corporate partners since only one used the lottery incentive.

Enrollment Challenges: In-person vs Remote

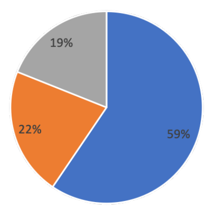
We faced significant challenges with the scale of our enrollment being both large and nationwide. In-person enrollments included a mix of scheduled participants and walk-ins, which we suspect increased our participant count at each enrollment site through word of mouth. In our largest in-person enrollment, we sent four researchers to a large corporate campus for 10 business days to enroll a cohort of 163 participants. This created a controlled environment and enabled researchers to troubleshoot and verify that everything was working, as well as allowing participants to engage in conversations about the consent form. However, we found that in-person enrollments incurred significant time and travel costs for the research team, as well as issues with scalability. Specifically, the rate of enrollment was constrained by the number of researchers able to administer survey instruments, the size of the room available at our organizational field sites, and the amount of equipment transported. In order to enroll a larger number of participants, we turned to a remote enrollment model to supplement the above efforts and mitigate the challenges of an in-person approach. However, this introduced new logistical issues. We needed to collect further personal and identifiable information, such as shipping addresses in order to mail equipment. The in-person protocol also had to be modified to incorporate remote proctors during survey-taking. Specifically, credentials to access the IGTB surveys and device setup instructions, such as how to pair the wearable to the phone agent, had to be modified to be self-administered. The enrollment instructions and study credentials were then password locked and sent in advance to the participant. For the remote enrollment sessions, we used Zoom (https://zoom.us/), a video-conferencing software, with the password provided by the proctor. This prevented participants from taking the survey before they could be monitored remotely. Once these initial hurdles were overcome, we were able to limit the researcher's time spent enrolling participants as single proctors could remotely monitor 10 participants simultaneously by assigning participants to private video chats using Zoom's breakout room functionality. Further, researchers from different time zones could cover more sessions, allowing participants to enroll within 8 am and 10 pm (local times). We had a combination of student and postdoc researcher proctors (6) and externally paid proctors (3) to conduct these remote enrollment sessions.



(a) In Person Direct Payment



(b) Remote Direct Payment



(c) Remote Lottery

■ Fully Compliant ■ Mostly Compliant ■ Non Compliant

Figure 1: A comparison of compliance between Direct and Lottery payment methods.

However, remote enrollment had its downsides, including a wide range of technological failures, e.g. incompatible browsers, disconnections, non-functioning webcams or microphones. Further, many participants chose to enroll from home, and some were reluctant to video conference from private locations. In addition, participant-chosen remote enrollment environments had more distractions (e.g. from pets or family members) than in-person enrollments. Once participants had unlocked the instructions, they sometimes missed steps, failed to read/follow instructions, or faced common technical difficulties without a researcher present to immediately troubleshoot. This necessitated time-intensive followups via email, scheduled Zoom troubleshooting sessions, and an online portal where participants reported bugs and other issues. Last, remote enrollment required mailing equipment to participants without assurance that they would ultimately enroll in the study. 37 kits were sent to participants who never enrolled in the study, in spite of numerous follow-ups via email, text, and phone. Efforts to retrieve these kits by sending pre-paid return shipping labels returned only 7 kits.

COMPLIANCE

Enrollment occurred from January 2018 to July 2018. We have seen encouraging evidence that our study methods were effective in maintaining participant engagement and ensuring compliance. Thus far, we have had only 38 drops (approximately 5% of our participants). Our overall compliance score (phone agent, wearable, and daily surveys for the first 56 days) is 85.4% with 500 participants having an 80% or higher compliance score across all streams, 136 being mostly compliant (> than 60%, e.g., 100% compliance for the daily surveys and phone agent but not for the wearable), and 82 being wholly non-compliant (<60%). For social media compliance, see [5]. Figure 1 in the sidebar shows compliance rates across the three groups. The graph demonstrates that in person enrolled direct payment participants are more compliant than remotely enrolled direct payment participants. Also, lottery payments are less effective, though again the difference could simply be attributed to differences in how compliance benchmarks were computed and differences among corporate partners (see above).

PARTICIPANT PRIVACY

Participant privacy was a key concern for this study, given the large amount and sensitive nature of the data being collected. To ensure transparency, we generated a 8-page technical specification form to accompany the shorter 5 page informed consent form. Whereas the informed consent form provided high-level descriptions of each sensing stream, the technical specifications listed what each device was capturing and how it would be secured and stored. We also asked participants to specifically consent to each required and optional sensing streams individually. We further included provisions to clarify any questions via email or face-to-face discussions. Each data stream, and the study overall, has built in privacy and data security measures as follows.

For the wearable, after signing up for a Garmin account, participants granted an authorization token, which we used for data collection via Open Authorization (OAuth). This method allowed us to collect data without collecting sensitive information such as Garmin usernames and passwords. For the beacons, Gimbal randomizes each MAC address during broadcasts. Gimbal's proprietary approach to beacon detection ensures that the beacons can only be tracked by authorized study devices (e.g. phones running our phone agent). Another privacy advantage of Bluetooth beacons over GPS is that locational context is relative instead of absolute. Simply put, we can use beacons to identify "home" and "work" without knowing where home and work are located on the map. For social media data privacy, see [5]. For the phone agent, participants were assigned a randomly generated study ID and password. Data was locally spooled on the phone and periodically sent to a staging server using a secure HTTPS connection. Data was then securely stored (via a trusted parser) into the archival back end database where it is stripped of the study ID and assigned a new randomized ID for further anonymization because study IDs are necessarily known by proctors when enrolling participants. Wherever possible, communications to and from the database server used HTTPS/SSH using well-issued certificates. In addition, data was stored using randomized identifiers with a protected mapping, and only de-identified data was used for analysis. Further, we divided front-facing and backend servers, such that incoming data is separated from archival data, limiting the negative impact of a potential breach of the front facing server.

DISCUSSION

Large-scale studies with multiple sensing streams allow researchers to understand nuanced patterns of people in their natural environments, with little intrusion, and across extended time frames. Given the considerable time, effort, and expense to recruit and enroll participants, it is important to retain as many participants as possible. To date, we retained 95% of our study participants over halfway through the study duration. We attribute this to our privacy efforts, our compensation schemes, the unobtrusive nature of the sensors, and the potential value of the study. Our study design emphasized privacy with security built into multiple levels of data collection, and we invested significant effort to ensure participants understood what was and was not being collected, how data would be used, and how it would be kept private and secure. In addition, our study was minimally intrusive, with an initial 2-hr setup and brief 5-minute surveys for the following 56 days. With the exception of the wearable, the other sensors were passive and did not require user input - for example, the home Bluetooth beacon just remained in place, the phone agent ran in the background while the phone was on, and social media data was collected in the background. Further, our data suggests that remote enrollment is a viable method for enrolling a geographically diverse set of participants for large-scale ecologically valid research, though it may come with less participant compliance. We have many exciting steps ahead. For one, we are working on ways to present data patterns to participants so they can use it to drive reflection and behavioral changes. We are computing further features associated with theoretical constructs (e.g., work/life balance; sleep hygiene) and are testing their predictive and explanatory power with respect to workplace performance, psychological traits (e.g., personality, affect), and physical characteristics, and how these features fluctuate across time. Finally, we are exploring how we can use these insights for interventions. For instance, participants identified as sleep-deprived could receive texts with recommendations on how to improve sleep quality. In general, we expect to generate considerable insights from this unprecedented dataset, to both advance science and technology, and to drive interventions aimed at improving quality of life and work.

ACKNOWLEDGEMENTS

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800007. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

REFERENCES

[1] Daniel J Buysse, Charles. F Reynolds, Timothy H Monk, Susan R Berman, and David J Kupfer. 1989. The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research. *Psychiatry Research* 28, 2 (1989), 193–213.

[2] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2014. Capturing the Mood: Facebook and Face-to-Face Encounters in the Workplace. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*. ACM Press, Baltimore, Maryland, USA, 1082–1094. https://doi.org/10.1145/2531602.2531673

[3] Philip M. Podsakoff, Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology* 88, 5 (2003), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

[4] Rachael Purta, Stephen Mattingly, Lixing Song, Omar Lizardo, David Hachen, Christian Poellabauer, and Aaron Striegel. 2016. Experiences Measuring Sleep and Physical Activity Patterns Across a Large College Cohort with Fitbits. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 28–35. https://doi.org/10.1145/2971763.2971767

[5] Koustuv Saha, Ayse Elvan Bayraktaraglu, Andrew Campbell, Nitesh V Chawla, Munmun De Choudhury, Sidney K D'Mello, Anind K Dey, Ge Gao, Julie M Gregg, Krithika Jagannath, Gloria Mark, Gonzalo J Martinez, Stephen M Mattingly, Edward Moskal, Anusha Sirigiri, Aaron Striegel, Koustuv Yoo, Dong Whind Saha, Anusha Sirigiri, and Aaron Striegel. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In *CHI Ext. Abstracts*. ACM.

[6] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D. Abowd, and Munmun De Choudhury. 2017. Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. *Proc. IMWUT* 1, 3 (Sept. 2017).

[7] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*. ACM Press, Seattle, Washington, 3–14. https://doi.org/10.1145/2632048.2632054