First-Gen Lens: Assessing Mental Health of First-Generation Students across Their First Year at College Using Mobile Sensing

WEICHEN WANG, Dartmouth College, USA
SUBIGYA NEPAL, Dartmouth College, USA
JEREMY F. HUCKINS, Dartmouth College, USA
LESSLEY HERNANDEZ, Dartmouth College, USA
VLADO VOJDANOVSKI, Dartmouth College, USA
DANTE MACK, Dartmouth College, USA
JANE PLOMP, Dartmouth College, USA
ARVIND PILLAI, Dartmouth College, USA
MIKIO OBUCHI, Dartmouth College, USA
ALEX DASILVA, Dartmouth College, USA
EILIS MURPHY, Dartmouth College, USA
ELIN HEDLUND, Dartmouth College, USA
COURTNEY ROGERS, Dartmouth College, USA
MEGHAN MEYER, Dartmouth College, USA
ANDREW CAMPBELL, Dartmouth College, USA

95

The transition from high school to college is a taxing time for young adults. New students arriving on campus navigate a myriad of challenges centered around adapting to new living situations, financial needs, academic pressures and social demands. First-year students need to gain new skills and strategies to cope with these new demands in order to make good decisions, ease their transition to independent living and ultimately succeed. In general, first-generation students are less prepared when they enter college in comparison to non-first-generation students. This presents additional challenges for first-generation students to overcome and be successful during their college years. We study first-year students through the lens of mobile phone sensing across their first year at college, including all academic terms and breaks. We collect longitudinal mobile sensing data for N=180 first-year college students, where 27 of the students are first-generation, representing 15% of

Authors' addresses: Weichen Wang, Dartmouth College, Computer Science, Hanover, NH, 03755, USA, weichen.wang.gr@dartmouth.edu; Subigya Nepal, Dartmouth College, Computer Science, Hanover, NH, 03755, USA; Jeremy F. Huckins, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Lessley Hernandez, Dartmouth College, Computer Science, Hanover, NH, 03755, USA; Vlado Vojdanovski, Dartmouth College, Computer Science, Hanover, NH, 03755, USA; Dante Mack, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Jane Plomp, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Arvind Pillai, Dartmouth College, Computer Science, Hanover, NH, 03755, USA; Mikio Obuchi, Dartmouth College, Computer Science, Hanover, NH, 03755, USA; Alex daSilva, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Eilis Murphy, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Elin Hedlund, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Courtney Rogers, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Meghan Meyer, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, 03755, USA; Andrew Campbell, Dartmouth College, Computer Science, Hanover, NH, 03755, USA.

2474-9567/2022/6-ART95 $15.00
https://doi.org/10.1145/3543194

the study cohort and representative of the number of first-generation students admitted each year at the study institution, Dartmouth College. We discuss risk factors, behavioral patterns and mental health of first-generation and non-first-generation students. We propose a deep learning model that accurately predicts the mental health of first-generation students by taking into account important distinguishing behavioral factors of first-generation students. Our study, which uses the StudentLife app, offers data-informed insights that could be used to identify struggling students and provide new forms of phone-based interventions with the goal of keeping students on track.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Applied computing** → *Health informatics.*

Additional Key Words and Phrases: mobile sensing, mental health, first-generation students, first year

ACM Reference Format:
Weichen Wang, Subigya Nepal, Jeremy F. Huckins, Lessley Hernandez, Vlado Vojdanovski, Dante Mack, Jane Plomp, Arvind Pillai, Mikio Obuchi, Alex daSilva, Eilis Murphy, Elin Hedlund, Courtney Rogers, Meghan Meyer, and Andrew Campbell. 2022. First-Gen Lens: Assessing Mental Health of First-Generation Students across Their First Year at College Using Mobile Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 95 (June 2022), 32 pages. https://doi.org/10.1145/3543194

1 INTRODUCTION

The transition from high school through the first year at college is particularly difficult for students [11, 12, 17]. While many students quickly adapt and excel during their first year navigating various challenges on campus, other first-years experience increased stress, isolation, loneliness, anxiety and depression. In a survey of more than 150,000 students, 9.5% of the first-year students report feeling frequently depressed, whereas 34.6% report being overwhelmed by academic pressures and other demands [22]. It is important that students know how to address challenges and risks as they start their first year to grow as healthy young adults and succeed both socially and academically. Among all first-year students entering college, "first-generation" college students are one group that experience increased risk. These are students who come from families with no history of college degrees. Therefore, first-generation students have no family history of how to deal with the risks and demands of challenging academic environments, how to fit in, and how to cope with the various challenges they will encounter [21, 53, 67]. The risk of attrition in the first year among first-generation students is 71% higher than that for non-first-generation students [40]. Compared to their non-first-generation peers, first-generation students are more likely to face challenges that jeopardize their abilities to adjust to college life and achieve academic success [36].

Researchers have identified various risk factors associated with first-generation students, including, physical and mental health, family support in decision-making [93], social support and socioeconomic status (SES) [13, 93]. The higher education literature uses *at-risk* as a term for students who are poorly equipped to deal with challenges at college [36]. In [35], the authors discuss how first-generation students do not receive the same level of social support from their families and friends as their non-first-generation counterparts. The authors also report that first-generation students are more likely to experience depressive symptoms in comparison to non-first-generation students. First-generation students are also known to spend less time socializing with peers and interacting with teachers [78] which limits their network. Prior research also finds that students from lower socioeconomic backgrounds struggle with fitting in at college. Getting help, assessing options and planning for college are all novel challenges for them, making the transition to college even more difficult [19]. Given the enormous challenges facing first-generation students, it is not surprising to learn that they experience higher attrition rates [6, 56] and poorer mental health [41] in comparison to non-first-generation students.

While there is a wealth of literature demonstrating the need for educators and researchers to pay attention to the mental health of first-generation college students, there is no work that tracks and forecasts the mental health of this at-risk group. Beyond surveys, self-reports and anecdotal demographic descriptions concerning the stresses and strains of first-generation college life, we use a passive sensing tool and predictive analytics to offer insights into the mental health (e.g., depression, anxiety) of first-generation students. Specifically, we design a low-cost mental health prediction system for students at risk that aims to address the following research questions:

- Researchers have reported various risk factors associated with first-generation students. We aim to replicate these findings using mobile sensing technology.
  ***(Q1)** Are the first-generation students in our cohort at-risk? Do such risk factors have an association with their mental health during their first year at college? And, if so, how? More importantly, which of these risk factors might be most associated with mental health?*
- Given these risk factors, we hypothesize that first-generation students exhibit behavioral patterns distinct from non-first-generation students in the first year.
  ***(Q2)** What are the key behavioral differences between first-generation and non-first-generation students across each term as they progress through their first year? And how are these behaviors associated with risk factors?*
- We hypothesize that some first-generation students learn new coping skills to overcome barriers and challenges [21, 66] during their first year. As a result, they are less at-risk, enabling them to better transition to college and manage their mental well-being effectively like many non-first-generation students.
  ***(Q3)** What behaviors do first-generation students exhibit that are associated with better mental health? In particular, are the behaviors of these students that cope better similar to the behaviors exhibited by non-first-generation students?*
- The insights and outcomes from addressing the prior questions offer an opportunity to model and predict mental health from sensor data. One modeling challenge is overcoming the in-balance and possible bias between the minority population (i.e., 27 first-generation students) and the majority population (i.e., 153 non-first-generation students) in the modeling cohort.
  ***(Q4)** Can we accurately predict the mental health of the first-generation students using deep learning by taking into account important distinguishing behavioral factors of first-generation students?*

In recent years, passive sensing using mobile smartphone technology has enabled users to assess daily behaviors without burdening the user. For example, the StudentLife study [84] established the first link between passively sensed activities and mental health outcomes for college students. However, a common shortcoming of existing research is that it focuses on general enrolled students without considering the distinctive characteristics of certain groups, such as first-generation students. In addition, prior studies of college students have not considered the full first year of college, which is a critical period of transition for all students but particularly for first-generation students. As a result, the field is unable to gain an in-depth understanding of first-generation students' behavior and mental health across their full first year at college. In this paper, we study N=180 first-year students using mobile phone sensing throughout their first year. Notably, 15% of the 180 first-year students (N=27) are first-generation students. To control for selection bias as much as possible, the study was advertised to all first-year students during the enrollment period at Dartmouth College irrespective of whether they were first-generation students or not; that is, we avoid explicitly selecting only first-generation students. Even with such selection criteria in place, our First-Gen study population is proportionally representative of the number of first-generation students admitted to Dartmouth College on an annual basis; in fact, we have 15% first-generation students in our cohort of students, which is slightly higher than the normal rate admitted to the university (which varies between 10-12 % annually).

To the best of our knowledge, this is the first study to investigate the first-year student experience using smartphone data across a full 12-month period, including all their academic terms and their academic breaks when students typically return home. The longitudinal nature of the First-Gen study offers an in-depth portrait of the first year of college life. It allows us to explore the behavioral patterns and differences at a level of detail not possible before. Furthermore, it presents an opportunity to study the predictive nature of time series sensing data and its relationship with the mental well-being of first-generation students. In addressing the research questions discussed above, we make the following contributions:

- We capture and quantify the high school years as first-year students enter Dartmouth College using a high school life survey. We assess the risk factors for all students (N=180), taking into account various dimensions of the survey, including, socioeconomics, lifestyle, and social and support networks. We also use periodic Ecological Momentary Assessment (EMA) to collect self-reported mental health data from students. We find that first-generation students are at more risk based on their socioeconomic status, lifestyle and support network, but not in the area of sociability. Among these risk factors, however, lifestyle and sociability in high school (which are more about behavioral patterns than demographics) had a strong association with mental health. Lifestyle and sociability can be studied through passive-sensing behavioral data [32, 84]. This suggests that passive sensing may be beneficial in assessing mental health and that there might be some utility in exploring the use of passive sensing to understand the behavior of students.
- We use mobile sensing from students' phones to capture the behavior of first-generation and non-first-generation students during their first year. There are significant differences in behavioral patterns between these two groups. For example, first-generation students spend more time in study areas and seldom visit the gym and Greek houses where campus-wide social events and parties usually occur. First-generation students also exhibit different behaviors than their non-first-generation counterparts. For example, they experience less regular sleep patterns and are more regular in the places they visit on campus.
- As each term of the academic year progresses, we capture changes in inferred behaviors associated with better mental health for first-generation students as they adapt to overcome initial risk factors. We find that such behaviors are not identical to the behaviors of the non-first-generation students; that is, it appears as though first-generation students have unique correlations between behavior captured by smartphone sensing and mental health. For example, a longer phone unlock duration at study places may indicate a deterioration of mental health for first-generation students, whereas spending less time at Greek houses and having a less regular lifestyle are strong indicators of poor mental health for non-first-generation students,.
- Based on the observations discussed above, we design a novel deep learning model for mobile sensing time series data that pays attention to first-generation students when predicting their mental health. We propose a new deep model architecture that improves the overall F1-score by 0.07 – increasing it from the baseline of 0.63 to 0.70. More importantly, we discover that learning models based on generic architectures found in the existing literature are biased towards the majority (i.e., non-first-generation students) and perform poorly on the minority population (i.e, first-generation students). Our First-Gen deep learning architecture removes this bias and improves the F1-score for first-generation students from 0.58 to 0.71. We can boost the performance of our predictive model by capturing the important behavioral differences between the two student groups.

The structure of the paper is as follows: We first discuss related work in Section 2, followed by visualization and initial analysis of first-year students in Section 3. Following this, we present the details of our First-Gen study in terms of study design, surveys, EMA design and feature extraction. As discussed above, we present our results in Section 5 and discuss how they support the four exploratory research questions that drive the study. In Section 6, we discuss our key research findings, implications of the study on college administrations and limitations of the work. We finish with some concluding remarks in Section 7.

2 RELATED WORK

The start of college life represents one of the most significant transitions in a young person's life. Fromme et al. [29] conducted an online survey of N=2025 students and investigated behavioral risks during the transition to college from high school life using six metrics; these are, alcohol use, drinking and driving, aggression, drug use, crime and the number of sexual partners. In general, the authors [29] find alcohol use, marijuana use and sexual relations increase due to newly acquired freedom by the students. In another study of N=1453 students, the authors [11] report that the performance on standardized tests (e.g., the SAT) predicts mental health in the transition period from high school to college. Some researches imply that higher peer ability (i.e., having more "advantaged" classmates who come from families with high levels of parental education and income [33]) has a detrimental impact on student outcomes such as educational aspirations. Self-concept theory, sometimes known as the "big fish, little pond" effect, is a classic psychological theory that explains how peer ability and the person's place within the group's ability distribution impacts individual learning [51]. As peer ability increases after high school, students' sense of worth reduces due to the relative decline in their academic performance. This is akin to feeling like a "big fish in a small pond" and the sense by students that all "fish" is big at college. In addition, the first term of the first year plays an important role in the academic year. For example, many students are unable to reach their target GPA established at the beginning of their first term [42]. This phenomenon is partly due to students devoting a considerable amount of time during their first term to socializing and expending capital to establish their social network rather than dedicating time to studying and advancing their academic goals. As a result, students typically spend more time on academics during their second term onward [79].

There is considerable research on using mobile sensing to infer human behavior across various fields. In terms of passive sensing and mental health, many studies have emerged [31, 60, 61, 84, 90] reporting on depression [71, 87], anxiety [9, 65], stress [7, 62] and mood [47, 95]. The StudentLife study [84] first reported on how passively sensed behaviors from phones, such as conversational interaction, sleep and activity are associated with mental health outcomes and academic performance for 48 college students enrolled in one class over a single term. In another study of N=117 students living in a university's dormitory, researchers [7] trained a Random Forest machine learning model based on the daily weather, personality of the students and smartphone sensing data to classify the stress level of students with an accuracy of 72%. Researchers also discovered growing evidence that mobility features relate to depression [15, 70, 71]. Finally, Xu et al. [92] conducted a study of N=188 undergraduate students and captured routine behaviors, as well as behavior pattern differences between depressive and non-depressive subgroups.

3 FIRST YEAR IN A NUTSHELL

In this section, we present an overview of the students included in the First-Gen study (N=180) across their entire first year at Dartmouth College. We present a set of time series visualizations of inferred student behaviors and self-reports representing their first year in a nutshell.

3.1 College Life

Dartmouth College is a highly competitive academic institution located in a small college town in the northeast United States. Unlike universities in big cities, Dartmouth College is located on a self-contained campus. All first-year students are required to live on-campus, making on-campus mobile sensing meaningful because the campus is the very place where all student activities happen. Almost all the on-campus buildings are associated with a primary function, such as dorms, classrooms, library, gym, cafeteria, partying/social, etc. The number of multi-functional tall buildings on campus is negligible. Therefore, even without knowing the precise indoor location, the broader GPS signal indicates the primary type of area the students are in.

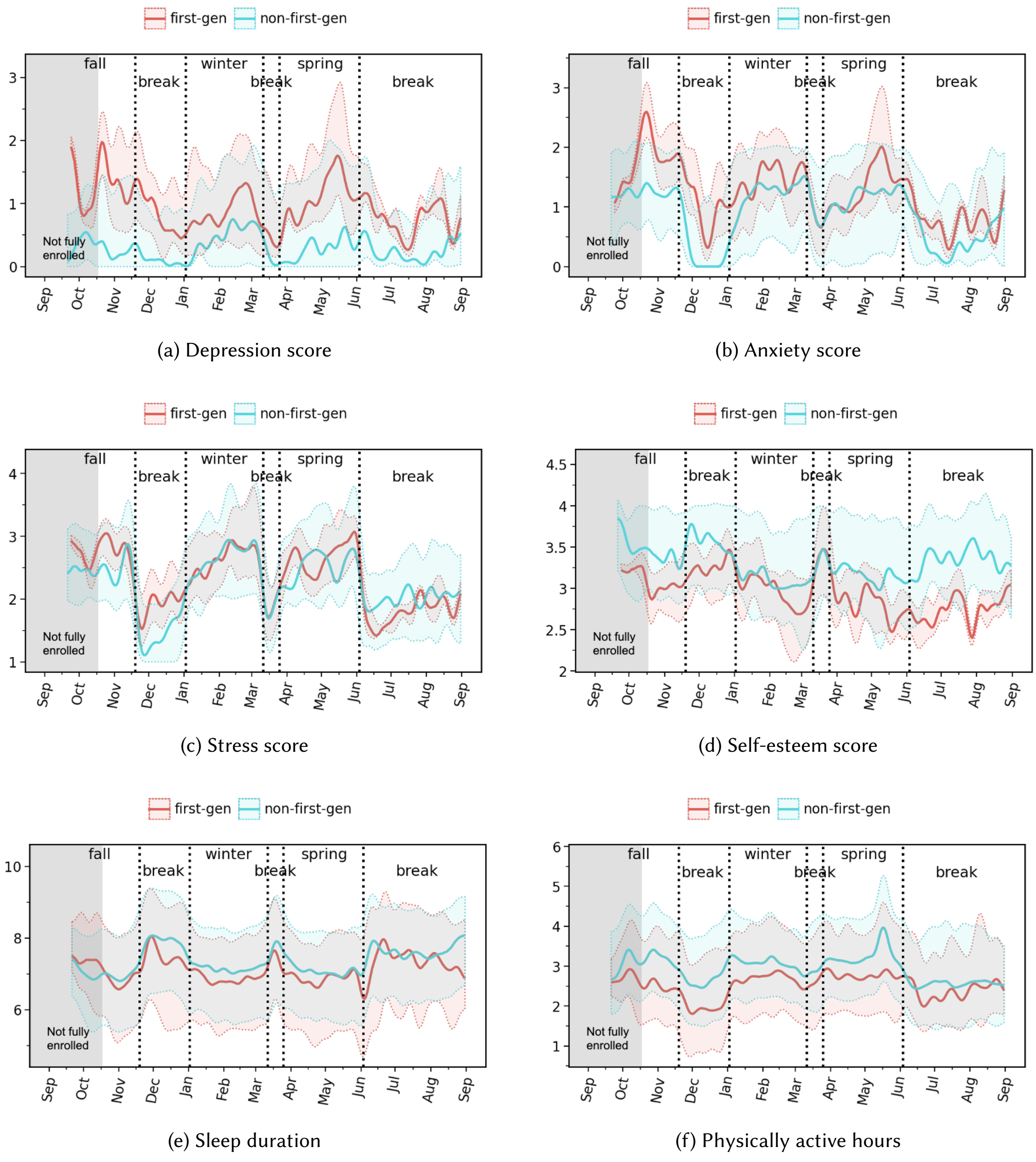
There are three major terms during the academic year at Dartmouth, each lasting ten weeks: the fall term (mid-September to November), the winter term (January to early March) and the spring term (late March to early June). There are three academic breaks during the year (viz. winter break, spring break and summer break) when undergraduates leave campus and typically return home, go on vacation, or take up internships in their hometown or away from home. All the terms are fast-paced compared to a 15-week semester system. Thus, students face an intense midterm period soon after the term begins. The major is not fixed for incoming students. During the first year, students need to seek out where their interests lie and the major they will work on in the years ahead. Students are recommended to select three courses each term. The largest on-campus social week of the year occurs in mid-May to bring the entire community together.

3.2 First-Year Data

In what follows, we present trends in student behavior and mental well-being across each term and break for the incoming first-year students. Figure 1 shows the dynamics of mental health across the entire year for the cohort based on weekly self-reports and passively inferred behaviors (e.g., sleep duration, physical activity, etc.) obtained from students' phones. Collectively, these time series plots provide an interesting glimpse into the complex lives of first-year students in the First-Gen study. Note, due to the rolling enrollment process, the plots in the first half of the fall term (Sep-Oct) are based on incomplete participants and may not be as reliable on the aggregate as the later months.

Figure 1a to 1d show the fluctuation of mental health over the year as measured by Ecological Momentary Assessments (EMAs). In the fall term, the term immediately after enrollment, we can see that depression and anxiety increase gradually, whereas self-esteem rapidly declines. We observe that first-generation students appear to have poorer mental health, particularly during the first term, probably due to the difficulty of adjusting to a highly competitive college environment. During the first winter break after the end of the first term, self-esteem improves dramatically and anxiety and stress fall to their lowest levels. When students are home, we can observe that their mental state rebounds back to their baseline as they decompress, sleep, and relax away from academic and other pressures (e.g., social, sports, financial, relationships). After returning to campus for the winter term, self-esteem starts to decline again while mental health improves. While these trends are common for all the terms, the relationship between self-esteem and other metrics is reversed during breaks. Mental health improves during the spring break. However, during the spring term, depression, anxiety, and stress rise throughout the term. In particular, depression peaks in May, when self-esteem is at its lowest point. Throughout the entire year, we observe a pattern of mental health deteriorating as the term progresses and a bounce back during breaks. However, the degree of deterioration during the term is larger than the degree of recovery during holidays. As a result, mental health progressively deteriorates through the year. At the start of the longest break of the year (i.e., the 3-month summer vacation), depression, anxiety and stress dip to the same level observed during spring break – back to the baseline. These mental health indicators start rising before the start of the new term representing a U-curve in the figure. We also observe the same anticipatory mental health response before the students in our study return for their second year.

Figure 1e and 1f show how much students are sleeping and being physically active across the entire academic year from mobile phone sensing. During the fall term, sleep duration significantly drops at the beginning of the term, rises, and then falls to its lowest point. Physical activity also seems to be affected by the academic calendar as it declines during the midterm exam period and toward finals. During winter break, sleep duration increases sharply from its lowest point to its highest peak. We presume students sleep more because there is no pressure from classes, tests, exams and social demands back home. Physical activity also drops to its lowest point during the winter break. We are likely observing seasonal influences during the winter term and break, as shown by a decrease in mobility and physical activity of the students in our study. Sleep duration during the winter term declines as the classes begin. We also observe a drop in sleep duration during the midterm exam period. Sleep duration and physical activity increase towards spring break. In mid-May, physical activity peaks during the year's biggest on-campus social event (i.e., the Green Key spring festival). This trend appears to be more pronounced among non-first-generation students in our study cohort.



(a) Depression score

(b) Anxiety score

(c) Stress score

(d) Self-esteem score

(e) Sleep duration

(f) Physically active hours

Fig. 1. Mental health dynamics, sleep and physical activity over the first year. Plots are based on the mean value and 25/75 percentiles of each measurement among students, grouped by first and non-first generations. Participants are completely enrolled by mid-October. The plot indicates that the first half of the fall term (Sep-Oct) has insufficient participation.

4 FIRST-GEN METHODOLOGY

In this section, we discuss the details of the First-Gen study, mobile sensing and self-report system and data. Later, in Section 5 and Section 6, we discuss our results and insights, respectively.

4.1 Study Design

The First-Gen study recruits N=180 first-year students, among which N=27 are first-generation students, representing 15% of the total participants. Note that the percentage of first-generation students admitted to Dartmouth College varies between 10-12% per year on average. Therefore, we have a greater first-generation sample than the typical incoming class in any particular year, which adds power to the study. This study is approved by the Institutional Review Board (IRB) at Dartmouth College. Students were recruited and consented to participate at the beginning of their first year when joining Dartmouth College from high school, starting in September 2017. Students who agree to participate in the First-Gen study install our data collection application on their Android or Apple smartphone. The mobile sensing application collects two types of data: sensing data to capture users' behavior and EMA to measure mental health. Participants were compensated for their weekly EMA responses at 10 dollars per week. The majority of our participants (67.8%, N=122) identify as female. In terms of race, 60% (N=108) are White, 23.9% (N=43) are Asians, 3.3% (N=6) are Black or African American, 2.8% (N=5) are American Indian/Alaska Native, 6.7% (N=12) belong to more than one race, and 3.3% (N=6) have not reported their race.

4.2 High School Survey

We use a high school survey to identify risks associated with students entering their fall term at Dartmouth College (Section 5.1). The survey, shown in Table 1, captures pre-college variables based on students' self-reported family background and experience during the high school years. Survey questions are selected by the psychologists in our research team from the widely adopted CIRP Freshmen Survey (TFS) [69] and focus on four major risk factors often included in educational literature [24, 25, 35, 36, 78, 83, 91, 93]: (1) socioeconomic status (SES) [83, 91], (2) lifestyle [24, 25], (3) support [35, 78, 93] and (4) sociability [78]. The original TFS is intended for use with incoming first-year students before their first day of classes. It has been administered to over 15 million students at over 1,900 institutions for over 50 years. Importantly, it has provided valid and reliable data on incoming college students' demographic characteristics, high school experiences, attitudes, behaviors, and expectations for college [76].

4.3 StudentLife: Behavioral Sensing App

4.3.1 App and System Design. We use the StudentLife mobile sensing app [84] for our First-Gen study. StudentLife has been used for several longitudinal studies in mental health across the United States [84–89] and it allows us to sense students' behavior using iOS and Android phones passively. We upgraded StudentLife to meet the demands of longitudinal studies. We also developed a dashboard for research assistants to monitor user compliance. During the study, 85.2% of Android and 87.3% of iOS phones collect more than 23 hours of data per day. If we change the threshold to 19 hours, the metric is 91.5% for Android phones and 93.8% for iOS phones. There is no significant difference in the data quality of the first-generation and non-first-generation students. Considering that few students turn off their phones during sleeping hours, we have one year of observations with high data coverage. At the end of their first year, 173 of the original 180 students completed the study, representing a retention rate of 96%, significantly higher than existing longitudinal mobile sensing studies listed in Table 2. Note that the other studies shown in Table 2 are for shorter duration periods. Those that provided monetary compensation, paid students at a similar or higher weekly rate than the First-Gen study (i.e., greater than $10 per week).

To promote high compliance across the study's entire year, we adopt the following app design, integration, and deployment strategies. First, we use the AppCenter [57] for the software development toolkit (SDK) management

Table 1. High School Survey based on the CIRP Freshmen Survey (TFS) [69]: Questions associated with risk factors.

Questions	Likert Scale Options (score 1-5)
Socioeconomic Status	
I perceive myself as:	Lower socioeconomic class, Lower-middle socioeconomic class, Middle socioeconomic class, Middle-upper socioeconomic class, Upper socioeconomic class
Lifestyle	
During your time in HIGH SCHOOL:	
How would you rate your physical well-being?	Very poor, Poor, Average, Good, Very good
How would you rate your mental well-being?	Very poor, Poor, Average, Good, Very good
How physically active were you?	Not at all active, Slight active, Somewhat active, Very active, Extremely active
Did you have healthy sleeping patterns?	Not at all, Slightly, Somewhat, Very, Extremely
Did you have healthy eating patterns?	Not at all, Slightly, Somewhat, Very, Extremely
Support	
During HIGH SCHOOL, how often did you feel?:	
That my family provided me with the support that helped me succeed.	Never, Seldom, Sometimes, Often, Almost always
That teachers provided me with feedback that helped me assess my progress in my classes.	Never, Seldom, Sometimes, Often, Almost always
Sociability	
During HIGH SCHOOL, how often did you feel?:	
Lonely or homesick	Never, Seldom, Sometimes, Often, Almost always
Isolated from school life	Never, Seldom, Sometimes, Often, Almost always
How satisfied were you with your social life?	Not at all, Slightly, Somewhat, Very, Extremely

and app deployment. Specifically, the App Center is a continuous integration and delivery platform for iOS and Android phones, enabling fast, convenient and low burden app development and release cycles for pushing app updates to student phones. The SDK helps to monitor software crashes on the phone and allows researchers to push an update to users after fixing problems immediately. Next, we design the iOS version of the StudentLife sensing app based on Voice over Internet Protocol (VoIP) push notifications. Continuous sensing on iOS has historically been challenging because of the restrictive nature of the environment. Because iOS does not support "real" multitasking, a third-party app, such as StudentLife, will often have a limited amount of time available to execute when the system switches it to the background mode [3]. Existing mobile sensing applications rely on various background task declarations to resolve this. For example, the AWARE platform [4] claims to be a navigation program that constantly updates users on their location with the GPS sensor always on. Another iOS issue that can limit sensing apps is that missing data can arise if the background process is canceled (e.g., when insufficient memory is available), making the app unable to resume normally. We redesigned the iOS sensing system based on Voice over Internet Protocol (VoIP) push alerts to address these issues. By enabling StudentLife to make phone calls over an internet connection instead of cellular service, we trigger the application's sensing duty cycle with the help of oncoming traffic (i.e., push notifications to the app). Typically, StudentLife is paused in the absence of such traffic. We send push notifications to StudentLife from the backend, allowing it to wake up at regular intervals to run its sensing and inference pipeline. This technique has two benefits: first, the app is suspended and consumes almost no energy when sensors are not used. Second, this technique considerably enhances the system's robustness; even if an undesirable crash occurs, the app will resume functioning without user intervention. Finally, we conducted two pilot studies and solicited feedback before the main First-Gen study. The first pilot consists of a two-week focus group with 10 undergraduate students. Everybody is required to report the battery usage of our app daily, which is displayed on the system's battery management page of the study dashboard. StudentLife consumes between 3% and 8% of the total energy. Almost all students claimed that

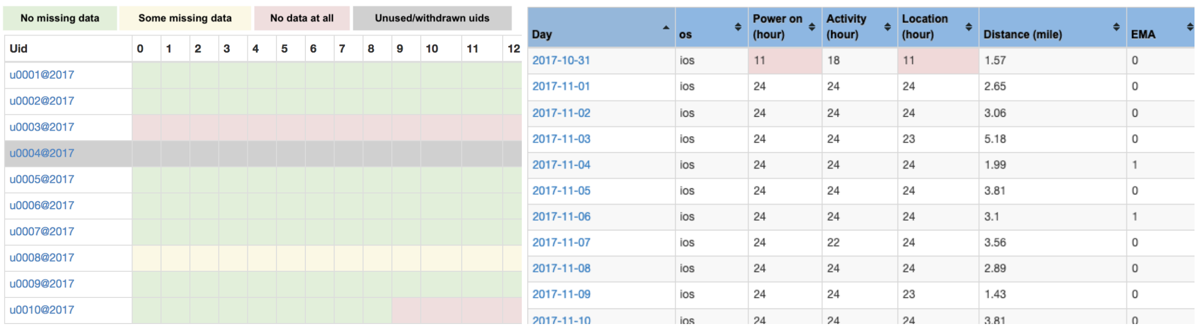
Table 2. Compliance in existing literature.

Reference	Target	Recruited participants	Duration	Compliance	compensation
Xu et al. 2019 [92]	depression	188 + 267 college students	106 days	Phase 1: 73% and Phase 2: 79%	A Fitbit Flex 2 ($100) and cash to $205 based on compliance
Wang et al. 2018(a) [88]	personality	646 college students	2 weeks	159 participants (24.6%) gave more than 7 days with 19+ hours of sensing data	Students did self-tracking using apps as part of a course assignment
Wang et al. 2018(b) [87]	depression	83 undergrads. students	9 weeks		Not revealed in paper
Boukhechba, 2018 [8]	mental health	72 undergrads. students	2 weeks	EMA compliance 72%	
Zhang, 2017 [95]	mood	42 college students	1 month	30 students (71.4%)	Not revealed in paper
Farhan, 2016 [27]	depression	79 college students	5 months	44.3% (GPS+PHQ9) 56.3% (Activity + PHQ9)	$15 Amazon gift card for every two weeks of active participation.
Huang, 2016 [37]	social anxiety	18 undergrads. students	10 days		Course credits and money (amount not revealed)
Sano, 2015 [73]	academic performance, sleep, stress, mental health	66 undergrads. students	30 days		Not revealed in paper
Wang, 2014 [84]	mental health, academic performance	60 college students	10 weeks	48 subjects (80%) completed	T-shirt. 10 Jawbone UPs and 10 Google Nexus 4 phones to top student collectors.

they did not alter their charging practices due to installing StudentLife on their phones. In addition, students indicated that they did not anticipate issues with using the app for a longitudinal one-year study. For the second pilot, we recruited 45 students for four-week testing period of the app and system. Several software bugs were fixed during this period, and a 6% average battery usage was reported during the trial.

StudentLife is designed to operate in the background of phones and passively collect sensing data without any user intervention. This design feature enables the app to sit "quietly" on phones and, importantly, makes large-scale, low-burden longitudinal studies feasible. Rather than collecting raw accelerometer data, we use the iOS Core Motion API and Google Activity Recognition API to record users' physical activities (e.g., stationary, walking, running, cycling, in a car) to conserve energy and enhance inference quality and accuracy. StudentLife collects raw data on phone usage and location, phone locks/unlocks and GPS locations every 10 minutes. This data is saved locally on the phone. When the phone detects an internet connection, StudentLife uploads the stored sensing to secure servers before wiping the data from the phone. The AWS server processes the raw data and calculates the behavioral attributes discussed in the next subsection. The StudentLife system also includes a dashboard (as shown in Figure 2) used to quickly inspect and visualize the data collected from users, as well as port EMA and overall compliance rates. The dashboard is used daily by researchers responsible for "data sitting" the study to identify problems proactively. Figure 2 shows compliance data displayed in the dashboard. Figure 2a shows the last 30-days (only 12 days are displayed) of compliance, where each row corresponds to a participant. The page employs different colors to indicate different compliance rates (green indicates no missing data, yellow indicates some missing data, and red indicates no data), allowing us to quickly identify students in our study who are experiencing data collection issues and contact them to resolve the problem. If a certain

UID (user identification) as shown in Figure 2a is "clicked", the screen transitions to Figure 2b, which displays daily summaries of an individual student's data quality. This includes the number of hours the phone is powered on, the number of hours of collected activity labels, the number of hours of the GPS data available, the distance traveled, the number of EMA the user completed, and so forth.



(a) Last 30-day compliance visualization. (b) Daily summary history of a particular user.

Fig. 2. The StudentLife dashboard used to "datasit" the First-Gen study.

4.3.2 Feature Extraction. We generate several features from the collected mobile sensing data. In what follows, we provide a high-level overview of these features; for a full list of behavioral features, see Table 3.

Physical Activity. Physical activity has been found to be associated with mental well-being. The MONARCA research [64] initially reported on data from mobile sensing and bipolar disorder, examining relationships between activity levels throughout the day and mental evaluation scores linked with the depressive spectrum. The findings in [1] suggest that circadian stability is a measure that can aid in the effective management of bipolar disorder. We use the iOS Core Motion API [39], and Google Activity Recognition API [30] to determine the physical activity of a student.

Phone Usage. Researchers [71] identified correlations between phone usage features from mobile phones and the severity of depressive symptoms, as measured by the 9-question Patient Health Questionnaire (PHQ-9) [43] in 40 subjects. PHQ-9 is a diagnostic tool used to screen for the presence and severity of depression. It rates depression based on the self-reported Patient Health Questionnaire. Previous research [20] indicates excessive smartphone usage is connected to depression or anxiety. In [38], researchers identified a positive relationship between smartphone screen time (e.g., phone unlock duration) and resting-state functional connectivity (RSFC) between brain regions associated with depression and antidepressant treatment response. The StudentLife app tracks the number of phone locks and unlocks students perform. We calculate both the total number of phone locks and unlocks, and the average time between phone locks and unlocks.

Mobility and Semantic Locations. Researchers discovered that mobility features are associated with depression. Wang et al. [87] found that mobility and location features act as proxy measures of decreasing interest or pleasure in activities. StudentLife sample GPS every 10 minutes to balance energy conservation and data quality. Raw GPS coordinates are first clustered using density-based spatial clustering with noise (DBSCAN) [26]. Following this, we calculate the number of unique locations and the distance traveled. In the First-Gen study, we use a campus-wide map of buildings to categorize the semantics of locations, such as study areas, dorms, social spaces, gyms, Greek houses, and so on, based on their primary function to understand better how much time students spend at the various semantic locations across campus. Based on this semantic understanding of locations, we compute contextually aware behavioral features; for example, we can learn how long students use their phones (lock/unlock) while in study areas, dorms, etc. We also compute class attendance based on students' registered courses and classrooms provided by the registrar's office. This strategy has been validated during several prior studies [84, 85, 87] at Dartmouth College.

Sleep. Sleep changes are one of the common symptoms associated with major depressive disorders [23]. Demirci et al.[20] discovered the association between sleep quality, depression and anxiety in 319 university students. We infer sleep duration, bedtime, and wake-up time using the method described in [16, 84], which had an accuracy of +/- 32 minutes to the ground truth.

Regularity. Additionally, we calculate the circadian rhythm in physical activities using the method described in [88]. We also compute location regularity. Previous studies [15, 55, 88] calculate location regularity based on whether participants were at the same GPS coordinates during the same period across different days. On the other hand, campus life has a distinct week-to-week rhythm, and buildings with similar primary functions are dispersed throughout campus. As a result, we compute the regularity of the locations using the Levenshtein distance [46] of the location categories across the same days on different weeks. We also compute the mean squares successive difference [82] *(MSSD) of sleep duration and bedtime.* MSSD measures the degree of autocorrelation and serves as an indicator of stability for things such as sleep. Regularity indexes are commonly explored in mental health sensing studies, and many of them report finding a significant association between the two factors [15, 63].

Table 3. Features computed from StudentLife mobile sensing data. Daily and epoch-based sensing features are computed across meaningful daily epochs: night/morning (12am-9am), day (9am-6pm), and evening (6pm-12am), allowing analysis of behavioral features and trends across different periods of the 24-hour clock).

category	**details**
physical activities	duration walking / running / cycling / in vehicle / sedentary
mobility and semantic locations	number of locations visited, distance travelled, max distance from the center of campus, duration at own dorms, other's dorms, food area, gyms, study places, social places(Greek houses), class attendance.
phone usage	number of lock/unlocks & unlocked duration at all places, study places, social places, own dorms, and other's dorms,
sleep patterns	sleep duration, sleep start time, sleep end time
regularity	location regularity (all categories, food and eating related, home related), circadian rhythm in physical activities, regularity of sleep duration, sleep start time, sleep end time over the past week.

4.4 Ecological Momentary Assessment (EMA): Self-reported Mental Health

We track the dynamics of student mental health using a built-in mobile EMA component integrated into the StudentLife app. Self-reports are randomly delivered to each student's phone once every week. We landed on this as a period that could scale across a year and not be considered burdensome to users. We use a PHQ-4 [44] to evaluate depression and anxiety once per week. PHQ-4 asks questions about the frequency of experiencing "feeling nervous, anxious or on edge", "not being able to stop or control worrying", "feeling down, depressed or hopeless" and "little interest or pleasure in doing things" over the last two weeks respectively, all using a 4-point scale: 0: not at all, 1: several days, 2: more than half the days, 3: nearly every day. We collected 4942 EMAs from 180 students over the entire year of the First-Gen study.

5 ANALYSIS

In this section, we discuss the results of the First-Gen study. Specifically, we address each of the driving research questions in turn (viz. Q1-Q4) as discussed earlier and shown in Figure 3.

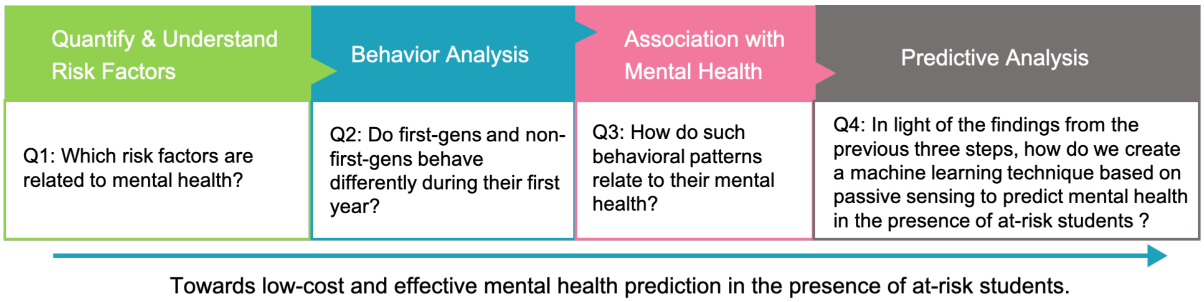


Fig. 3. Flow of analysis of research questions addressed by the First-Gen study

5.1 Q1: How Do the Risk Factors Associated with First-generation Students Relate to Their Mental Health during Their First Year at College?

Researchers have proposed various risk factors associated with first-generation students. In what follows, we break down these known risk factors into four pre-college metrics, as discussed earlier: socioeconomic status (SES), lifestyle, sociability and support. We quantify these risk factors and show how they relate to mental health. To do this, we take the following approach: (1) we compare risk factor scores between first-generation and non-first-generation students using a t-test; (2) we compare the mental health (using PHQ-4) of first-generation and non-first-generation students for each term using mixed-effects models; and finally (3) we analyze how these pre-college risk factors are associated with mental health. This helps us better understand the challenges faced by first-generation students during their first year. We perform various hypothesis tests in the following sections. To make sure we identify actual relationships and not spurious ones, we create two stratified, randomly sampled, non-overlapping subsets of students (each with N=90 students) while maintaining a similar ratio of first-generation to non-first-generation students in both subsets – 77 non-first-generations and 13 first-generation in Subset A, and 76 non-first-generations and 14 first-generation in Subset B. Throughout the rest of the paper, we refer to subset A using the symbol Ⓐ and subset B using the symbol Ⓑ. We then test our findings in these subsets and verify whether they hold true for each subset. Furthermore, we use a two-stage Benjamini-Hochberg method (TSBH) [5] to calculate the false discovery rate (FDR), ensuring a rigorous approach to analysis.

5.1.1 *Comparing Risk Factors between First-gens and Non-first-gens.*

We discuss four risk factors collected from our high school survey. Table 4 shows the average score and standard deviation of each risk factor described in Table 1. A lower score on an item indicates that students in a particular group (viz. first-gens and non-first-gens) are *at-risk* for that factor. The scores are normally distributed ($p > 0.05$ in the Kolmogorov Smirnov normality test [48]). We compare the scores for each risk factor between first-generation and non-first-generation students using a t-test. Outlier points greater than (mean + 2 * std) or less than (mean - 2 * std) are eliminated before the test to avoid producing erroneous statistical results. The results show that the first-generation students have significantly lower socioeconomic status (SES), poorer physical and mental health (referred to as lifestyle in the table), and are less satisfied with support from their families and teachers during high school. Note that these findings hold true for both the subsets A and B of students. However, there is no significant difference in social activities between the first-gens and non-first-gens during high school. This helps us better understand the real differences in at-risk factors between first-gens and non-first-gens.

Table 4. Comparison of risk factors between first-generation and non-first-generation students using T-test. The scores of risk factors are normally distributed (p > 0.05 using KS-test). Full statistical results on subset A and subset B can be found in the supplementary document.

risk factors	first-gen mean(std)	non-first-gen mean(std)	t-test p-value	significance holds true on subsets
SES	1.70(0.72)	3.72(0.91)	< **0.001***	Ⓐ Ⓑ
Lifestyle	2.87(0.64)	3.55(0.74)	< **0.001***	Ⓐ Ⓑ
Sociability	3.67(0.72)	3.81(0.67)	0.35	
Support	3.53(0.87)	4.13(0.64)	< **0.001***	Ⓐ Ⓑ
* p-value <= 0.05; ** p <= 0.01; *** p <= 0.001, bold if FDR-adjusted p <= 0.05				

5.1.2 *Comparing Mental Health of First-gens and Non-first-gens.*

Next, we investigate whether the challenges first-generation students face as they adapt to college life influences their mental health. Specifically, we compare mental health (using PHQ-4) across each term. PHQ-4 is used clinically to assess depression and anxiety. We use a contrast variable to indicate first-generation students and non-first-generation students. Group differences are examined using linear regression models [50] with the contrast variable as a predictor. Given the repeated measures design of the PHQ-4 data, we use mixed effects models with repeated PHQ-4 scores nested within an individual, and the group contrast is treated as the fixed effect.

Table 5. Comparison of PHQ-4 scores between first-generation and non-first-generation students using mixed effects with repeated PHQ-4 scores nested within the individual, and the group contrast (1: 1st-gens & 0: non-first-gens) treated as a fixed effect. First-generation students had significantly higher depression scores during the fall term only. The coefficients β decrease as the year progresses, possibly indicating that first-generation students adapt to overcome initial risk factors. Full statistical results on subset A and subset B can be found in the supplementary document.

term	first-gen mean(std)	non-first-gen mean(std)	mixed effects result	significance holds true on subsets
fall	3.29(2.12)	2.26(1.74)	$\boldsymbol{\beta = 1.12, SE = 0.39, p = 0.004^{**}}$	Ⓐ Ⓑ
winter	2.51(2.31)	2.21(1.97)	$\beta = 0.45, SE = 0.36, p = 0.21$	
spring	2.38(2.25)	2.30(1.99)	$\beta = 0.30, SE = 0.45, p = 0.50$	
* p-value <= 0.05; ** p <= 0.01; *** p <= 0.001; bold if FDR-adjusted p <= 0.05				

Table 5 shows the relationship between PHQ-4 scores and the two student groups (i.e., first-gen or non-first-gen). Note that, for the fall term, we only use EMA data collected during mid-term when all the participants had been fully enrolled in the study. The mean and standard deviation of the PHQ-4 score (averaged participant-wise) for both the first-generation and non-first-generation students are also included to provide additional insight into the data distribution. We report the coefficient (β), the standard error of the estimated coefficient (SE), and its significance as a result of mixed effects models. First-generation students have a higher PHQ-4 score (i.e., lower mental health status) than non-first-generation students. However, we observe that first-generation students' higher PHQ-4 scores are statistically significant only in the fall term ($\beta = 1.12, p = 0.004$), and the differences in PHQ-4 scores between the two groups of students during the winter and spring terms is no longer significant. The finding is valid on both subsets A and B. This result supports our hypothesis that the first-generation students adapt to overcome initial risk factors as the year progresses.

5.1.3 Examining the Correlation between Risk Factors and Mental Health. We examine how pre-college risk factors are associated with mental health. PHQ-4 scores are analyzed using mixed effects models with the repeated PHQ-4 nested within an individual, and each risk factor is treated as a fixed effect. As shown in Table 6, we find that higher SES is not significantly associated with better mental health (i.e., a lower PHQ-4 score), especially as the year progresses ($\beta = -0.15, p = 0.18$ for the fall term and it continues to be insignificant throughout the other terms). Better support is also not significantly associated with better mental health. However, better lifestyle ($\beta = -0.71, p < 0.001$ for the fall term and continues to be significant beyond the fall term) and sociability ($\beta = -0.69, p < 0.001$ in the winter term and $\beta = -0.62, p = 0.001$ in the spring term) indicates better mental health during the first year at college. It is important to note that many of these findings hold true for both subsets we perform tests on.

Table 6. Associations between risk factors and mental health using mixed effects models with the repeated PHQ-4 nested within the individual, and each risk factor (Likert scale from 1 to 5) treated as a fixed effect. Better lifestyle and sociability are associated with better mental health during the first year at college (i.e., lower PHQ-4 score). Full statistical results on subset A and subset B are available in the supplementary document.

risk factors	term	association with PHQ-4 based on mixed effects models	significance hold true on subsets
	fall	$\beta = -0.15, SE = 0.12, p = 0.18$	
SES	winter	$\beta = -0.01, SE = 0.11, p = 0.89$	
	spring	$\beta = 0.00, SE = 0.13, p = 0.99$	
	fall	$\boldsymbol{\beta = -0.71, SE = 0.16, p < 0.001}^{***}$	A B
Lifestyle	winter	$\boldsymbol{\beta = -0.84, SE = 0.14, p < 0.001}^{***}$	A B
	spring	$\boldsymbol{\beta = -0.89, SE = 0.17, p < 0.001}^{***}$	A B
	fall	$\boldsymbol{\beta = -0.51, SE = 0.17, p = 0.002}^{**}$	A
Sociability	winter	$\boldsymbol{\beta = -0.69, SE = 0.15, p < 0.001}^{***}$	A B
	spring	$\boldsymbol{\beta = -0.62, SE = 0.18, p = 0.001}^{***}$	A B
	fall	$\beta = -0.20, SE = 0.14, p = 0.16$	
Support	winter	$\beta = -0.20, SE = 0.13, p = 0.12$	
	spring	$\beta = -0.23, SE = 0.16, p = 0.15$	

* p-value <= 0.05; ** p <= 0.01; *** p <= 0.001, bold if FDR-adjusted p <= 0.05

5.2 Q2: What are the Key Behavioral Differences between First-generation and Non-first-generation Students across Each Term as They Progress through Their First Year?

In this section, we examine the differences in behavioral sensing data collected by smartphones. Behavioral sensing data and inferences are computed daily. We analyze the data using mixed effects models nested within an individual and treat the first-gens/non-first-gens group contrast as the fixed effect. For the fall term, statistical analysis is based on data collected after midterm, when participants are completely enrolled. In addition, features with strongly right-skewed distributions (e.g., time spent at Greek houses and time spent at the gym) are log-transformed, resulting in approximately normally distributed data.

Table 7 compares the behavior of first-generation and non-first-generation students based on mobile sensing data. We observe that first-generation students spend more time in the study areas than non-first-generation

Table 7. Comparison of behavioral features between first-generation and non-first-generation students. Behavioral sensing data and inferences are computed daily. They are analyzed using mixed effects models nested within individual and the first-gens/non-first-gens group contrast (1: 1st-gens & -1: non-first-gens) treated as the fixed effect. Associations between behavioral features and risk factors are examined using mixed effects models with the daily behavioral features nested within the individual. Each risk factor (Likert scale from 1 to 5) is treated as a fixed effect. The statistical analysis for the fall term is based on data collected after midterm, when participants were completely enrolled. Full statistical results on two subsets are available in the supplementary document.

behavior	first-gen mean(std)	non-first-gen mean(std)	mixed effects result	significance hold true on subsets	association with risk factors (p <= 0.05)
Duration at studying places (hr)					
fall	7.08(5.20)	3.71(3.66)	$\boldsymbol{\beta = 1.57, SE = 0.51, p = 0.002^{**}}$	Ⓐ Ⓑ	↓ SES
winter	6.28(5.22)	3.53(3.79)	$\boldsymbol{\beta = 1.40, SE = 0.47, p = 0.003^{**}}$	Ⓑ	↓ SES
spring	5.70(4.50)	3.02(3.41)	$\boldsymbol{\beta = 1.13, SE = 0.41, p = 0.006^{**}}$	Ⓑ	↓ SES
Class attendance rate					
fall	0.69(0.15)	0.69(0.17)	$\beta = 0.00, SE = 0.08, p = 0.978$		↑ lifestyle sociability support
winter	0.62(0.17)	0.59(0.18)	$\boldsymbol{\beta = 0.02, SE = 0.00, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ lifestyle sociability support ↓ SES
spring	0.62(0.18)	0.64(0.18)	$\boldsymbol{\beta = -0.02, SE = 0.00, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ lifestyle sociability support
Duration at food place (hr)					
fall	0.91(0.69)	1.21(0.64)	$\beta = -0.32, SE = 0.14, p = 0.02^{*}$	Ⓑ	↑ SES
winter	0.92(0.62)	1.05(0.65)	$\beta = -0.26, SE = 0.13, p < 0.06$		↑ SES
spring	0.84(0.59)	1.07(0.57)	$\beta = -0.13, SE = 0.12, p < 0.29$		↑ SES
Sleep duration regularity					
fall	-3.50(2.26)	-2.24(1.60)	$\boldsymbol{\beta = -1.00, SE = 0.06, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ SES lifestyle sociability support
winter	-2.33(1.97)	-2.35(1.16)	$\boldsymbol{\beta = -0.58, SE = 0.03, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ SES lifestyle sociability support
spring	-3.01(1.91)	-2.34(1.32)	$\boldsymbol{\beta = -0.63, SE = 0.03, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ SES lifestyle sociability support
Physical activity					
fall	2.57(1.2)	3.05(1.17)	$\beta = -0.35, SE = 0.25, p = 0.15$		↑ SES
winter	-2.72(1.24)	3.17(1.0)	$\beta = -0.3, SE = 0.22, p = 0.187$		↑ SES
spring	2.53(1.29)	3.37(1.06)	$\beta = -0.47, SE = 0.24, p = 0.05^{*}$		↑ SES
Duration at Greek house (log-transformed)					
fall	-20.14(3.93)	-17.97(3.42)	$\beta = -1.07, SE = 0.72, p = 0.13$		
winter	-19.63(2.47)	-17.58(2.74)	$\boldsymbol{\beta = -1.56, SE = 0.57, p = 0.006^{**}}$	Ⓑ	
spring	-19.80(2.10)	-17.45(3.26)	$\boldsymbol{\beta = -2.01, SE = 0.68, p = 0.003^{**}}$	Ⓐ	
Duration at gyms (log-transformed)					
fall	-20.46(3.74)	-17.60(5.02)	$\beta = -2.65, SE = 1.05, p = 0.01^{**}$	Ⓑ	↑ lifestyle
winter	-19.25(1.98)	-15.96(5.35)	$\boldsymbol{\beta = -4.20, SE = 1.05, p < 0.001^{***}}$	Ⓐ Ⓑ	↑ lifestyle
spring	-20.18(2.62)	-17.07(4.18)	$\boldsymbol{\beta = -2.69, SE = 0.89, p = 0.002^{**}}$	Ⓐ Ⓑ	↑ lifestyle

* p <= 0.05; ** p <= 0.01; *** p <= 0.001; bold if FDR-adjusted p <= 0.05.

: SES; : lifestyle; : sociability; : support; ↑: positive association; ↓: negative association;

students. This is consistent with the finding of [78], which indicates that first-generation students are more likely to use libraries and study spaces. Class attendance is an important metric when assessing college life. First-generation students appear to have a higher rate of winter class attendance than non-first-generation students, but a lower rate during spring classes. Prior studies suggest that such differences could be because of perceived classroom competitiveness, which contributes to differences in engagement, attendance and retention in courses [14]. While first-generation students seem to spend less time in the food area than non-first-generations in the fall, the difference diminishes and becomes insignificant during the winter and spring terms. First-generation students appear to have less consistent sleep duration than non-first-generation students. It could be that they are still adjusting to new academic demands, living arrangements, and social demands and thus end up cutting back on one of the few things they can control – the amount of sleep they get [18], resulting in poorer sleeping habits. First-generation students also spend less time partying at Greek houses. This may bolster the argument that first-generation students are less adept at cultivating social capital – privileged knowledge, resources, and information acquired through social networks – than non-first-generation students [10]. In addition, first-generation students appear to be slightly less physically active than non-first-generation students, although this difference is not significant in the fall and winter; they also spend less time at gyms. This is consistent with the finding in [59], which indicates that first-generation students are generally less physically active. Recall our finding from earlier that first-generation students are more at risk for lifestyle-related risk factors. Thus, first-generation students being less physically active ties back to such risk-proneness [24, 25]. Another possibility is that they are spending less time exercising due to other demands, such as working at the college, course load, surroundings and schedule [54]. Many first-generation students at Dartmouth work during the term on campus, doing various jobs to earn additional money to support living expenses.

Table 7 also illustrates the relationship between such distinguishing behaviors and risk factors, based on mixed effects models with the daily behavioral features nested within an individual and each risk factor (Likert scale from 1 to 5) treated as a fixed effect. We find that socioeconomic status (SES) is inversely associated with time spent at study locations on campus. Class attendance rate is positively related to a better lifestyle, sociability and support. The duration of time spent at food halls is positively associated with SES. Regularity in sleep duration is linked with improved scores across all four risk factors. Physical activity is positively related to SES. The duration of time spent at gyms relates to a healthier lifestyle. On the other hand, the duration spent at Greek houses is unrelated to any of the risk factors.

5.3 Q3: What Behaviors Do First-generation Students Exhibit That Are Associated with Better Mental Health? And Are These Behaviors Similar to Those Exhibited by the Non-first-generation Students?

To be able to answer this question, we build a LASSO [80] regression model to predict the mental health of both student groups using the sensing data. The LASSO (Least Absolute Shrinkage and Selection Operator) regression performs feature selection during the fitting process, excluding less important features concerning PHQ-4. We use the average score of PHQ-4 across a term for each participant as ground truth, indicating whether the subject is depressed or anxious during the term. Behavioral features are treated similarly, with all features aggregated into a term-long daily average for each participant. The behavioral features are then standardized across all subjects. Next, we analyze the coefficients of two linear models to see how each group's behavior relates to outcomes. For reliability, in Q3 and Q4, we skip the fall term from this analysis, where participants have different lengths of sensor data due to the rolling enrollment process. Table 8 shows coefficients of LASSO regression for estimating PHQ-4 of first-generation and non-first-generation students. Each variable is scaled to aid convergence and obtain regression coefficients that could be compared for relative importance.

For non-first-generation students, we observe very similar linear models for both the winter and spring terms, as the LASSO model selects common features with similar coefficients for both terms, as shown in Table 8. In the

Table 8. The coefficients of LASSO (Least Absolute Shrinkage and Selection Operator) regression for estimating PHQ-4.

	non-first-generation		first-generation	
	winter	spring	winter	spring
Time spent at Greek houses	-0.35	-0.57		-0.38
Time spent at own dorm	-0.19			-1.39
Time spent at other's dorm	0.12	0.12		
Time spent at study places			-1.11	1.17
Time spent at gyms	-0.21			
Number of location visited				-2.76
Physically active duration	0.51	0.68	-1.03	
Phone unlock duration at study places			1.53	0.28
Number of phone unlock			-1.07	
Sleep offset time				0.21
Sleep onset time		0.18		
Class attendance	-0.20	-0.42	-0.64	
Regularity of sleep onset	-0.14	-0.04		
Regularity of sleep duration		-0.29	0.09	
Circadian rhythm of physical activity	-0.44	-0.65		1.30
Eating regularity		-0.06	0.68	
Location regularity		-0.34	-0.91	-0.26

case of non-first-generation students, the regression model mostly relies on time spent at Greek houses, time spent at other students' dorms, physical activity, class attendance, regularity of sleep onset and circadian rhythm of physical activity. The time spent partying at Greek houses appears to be negatively associated with PHQ-4, indicating better mental health. Surprisingly, higher class attendance is associated with worse mental health. We also observe that regularity of sleep, physical activity, eating, and location routines appear negatively associated with PHQ-4. This indicates that the regularity of routines in life relates to better mental health.

Physically active duration (e.g., walking, running) has a positive standardized coefficient. This is an interesting observation that might lead us to falsely believe that more physical activity is associated with worse mental health, which goes against the findings of prior research [84, 86]. However, such coefficients can also be influenced by the correlation between features. When we check the correlation between variables, we find that for non-first-generation students, time spent at Greek houses is positively related to the physically active duration ($r=0.38$, $p<0.001$ in the winter term and $r=0.4$, $p<0.001$ in the spring term). Therefore, it appears that time spent at Greek houses may have absorbed some of the variances associated with the physically active duration. A deeper dive into the results (as shown in Table 8) shows that when LASSO does not select the time spent at Greek houses as a predictor, physically active duration has a negative standardized coefficient, thus indicating the more physically active first-generation students are the better their mental health.

Among first-generation students, we observe very distinct linear models across the winter and spring terms. In fact, there are only two features with a common direction shared by models during these terms; that is, the phone unlock duration at study places and location regularity between the same days of weeks. The phone unlock duration at study places has a positive coefficient, indicating that the increased phone usage at study areas relates to poorer mental health. This finding is in line with prior findings [86] where the authors hypothesize that phone usage in the classroom and study places is a potential indicator of a student's diminished ability to concentrate – one of the depressive disorder symptoms. Our analysis suggests that such a connection seems more obvious among first-generation students. Similar to the non-first-generation students, we observe location regularity is negatively associated with PHQ-4 for first-generation students. This might mean that a general regularity in semantic location routines between days possibly relates to better mental health. However, we observe some distinct patterns associated with first-generation students; for example, during the winter term, regularity of visits to food places has a positive standardized coefficient. Gathering at a common dining place may not just be related to the consumption of meals but other hidden activities such as sitting and chatting with friends [77], or in the case of many of our first-generation students working at cafeterias on campus. It may be possible that eating regularity may have a different contextual meaning for the first-generation students and non-first-generation students, especially at the early stages of their college life. We also observe that the time spent at Greek houses is selected as a predictor for the spring term only, possibly because the first-generation students adapt and participate in more social events at Greek houses as the year progresses (see Table 7). Visiting more locations on campus is associated with better mental health during the spring term but not during the winter term, possibly due to the seasonal effect (i.e., cold weather). Overall, we notice only a few commonalities in the linear models of first-generation and non-first-generation students. First-generation students have a range of unique behavioral predictors associated with their mental health. Such predictors appear to even change over time. As a result, first-gens start to overcome initial challenges, and by navigating various academic and social demands, they begin to adapt to college life positively.

5.4 Q4: How Can We Predict the Mental Well-being of At-risk Students More Accurately, with a Particular Focus on the First-generation Students?

Inferring mental health from passive sensing is not a novel idea, and machine learning has been widely used to accomplish this [81, 86]. Prior research has employed both classical machine learning [81, 86] and deep learning techniques [2, 75, 94]. However, previous research focuses exclusively on model performance over the entire participant population, without particularly accounting for a distinct subpopulation. Our exploratory research suggests that generic models found in the existing literature that are trained without taking into account the variations between first-generation and non-first-generation students are not optimal in assessing the mental health of first-generation students. However, training models individually between the two groups is not desirable because we lack adequate first-generation samples and because this strategy precludes the two groups from learning common information. As a result, we require an architecture that can be used to train the model on the complete dataset while also taking into account the distinctions between first-generation and non-first-generation students, as previously stated.

5.4.1 Data Preprocessing. The sensing data of one participant in a term can be viewed as a multivariate time series with a length equal to the number of days in a term. For each day in the term, we include the 83 features outlined in Section 4.3.2. Next, we standardize these features within and between participants. We employ robust scaling to deal with outliers, where the values of each variable are subtracted from their median and divided by the interquartile range (IQR) – the difference between the 75th and 25th percentiles. After handling outliers, we standardize features in two ways. First, we perform within-person standardization using all days in a term to capture intra-person variability in behaviors. Second, we standardize features across participants to capture inter-person or population-level information to improve model diversity. Ultimately, the two standardized feature sets are concatenated to form a time sequence with 166 features for each day. Each term is considered as a separate data point, with the dimensions (t, d), where t = 70 represents the number of days in a term and d = 166 represents the number of features. As discussed earlier, PHQ-4 is a two-week screening scale for anxiety and depression. We are interested in a student's mental health during the term. When a student has multiple EMAs within a term, researchers use the average of the responses during the term to indicate how the subject feels in general throughout the time period [38]. Thus, we use the average value of a student's PHQ-4 EMAs as the output ground truth. We use a cutoff score of 3 because clinically, PHQ-4 scores greater than 3 are seen in persons with anxiety and depression [44]. Using a cutoff score of 3, 34% of the data samples are positive. To ensure a rigorous evaluation, we set aside data from 20% of students (leave-subjects-out) for testing purposes only (we refer to this as D2) in the following sections. The remaining 80% of the data is used for training and validation (we refer to this as D1).

5.4.2 Baseline: Using Traditional Machine Learning Approaches. We first start with the basics and discuss whether traditional machine learning methods can predict students' mental health, considering first-generation students. Traditional machine learning approaches are incapable of directly handling time series. As a result, we take an average over the time axis and reduce the dimension of each data point to a vector of length d (where d equals 166, the number of features). Such a method is commonly used in current studies [81, 86]. In addition, two categorical features are added to the input: (1) whether a participant is first-generation, and (2) which term the data point belongs to. Next, we perform a grid search to optimize the model hyper-parameters within a stratified 4-fold leave-n-subject-out cross-validation using D1. We then train a model on the combined training and validation data using the selected super-parameters and evaluate this model on D2. Table 9 shows the predictive performance of several traditional machine learning classifiers. While these models can achieve precision scores slightly above 0.7 and an F1 score close to 0.7 across all participants, the predictive performance of these machine learning models for first-generation students is poor, as indicated by lower F1 scores. In addition, the recall rate is low. This could be because of the nature of such approaches. Traditional machine learning models cannot capture valuable temporal information after averaging over time. In addition, even if we indicate in the input that a student is a first-generation student, such algorithms will not fully exploit this information. For example, linear approaches may include an additional intercept on categorical variables (e.g., being first-generation). In contrast, decision tree-based methods may ignore such categorical features at higher levels if they cannot provide sufficient information gain. As a result, the model is more likely to be trained to fit the majority of students who are not first-generation.

Table 9. Predictive performance on PHQ-4 using traditional machine learning approaches. Traditional machine learning approaches fail to make predictions among first generations. Metrics are weighted to account for label imbalance.

Machine Learning method	precision (among all participants)	recall (among all participants)	F1-score (among all participants)	F1-score (among 1st-gen)
logistic regression	0.73	0.56	0.67	0.49
random forest	0.71	0.58	0.65	0.48
gradient boosting	0.73	0.56	0.69	0.42

5.4.3 Deep Learning Architecture. The aforementioned limitations of traditional machine learning models hinder the learning of good time series representations. In contrast, deep learning approaches can learn more sophisticated representations capturing the temporal nature of sensor data. Specifically, we propose a unique multi-task architecture that can leverage auxiliary information about the student's group (first-gen or not) and term (academic progress/ seasonal information).

According to Section 5.4.1, the dimension of each data input is (t=70, d=166), which represents a student's overall behavioral pattern over the duration of a term. We first use bidirectional Long Short-Term Memory (LSTM) network [34] to generate the representation of the input data. Bidirectional LSTM connects two hidden layers of opposite directions to improve model performance on sequence classification problems [74]. For each data point, we keep all hidden states at each step of the recurrent neural network so that the bidirectional LSTM layer generates an output H with the shape of (t = 70, 2 * hidden units). We further add self-attention [49], a technique widely adopted in natural language processing for sentence embedding, to the bidirectional LSTM models to assign different weights to different days in a term emphasizing their importance in predicting depression.

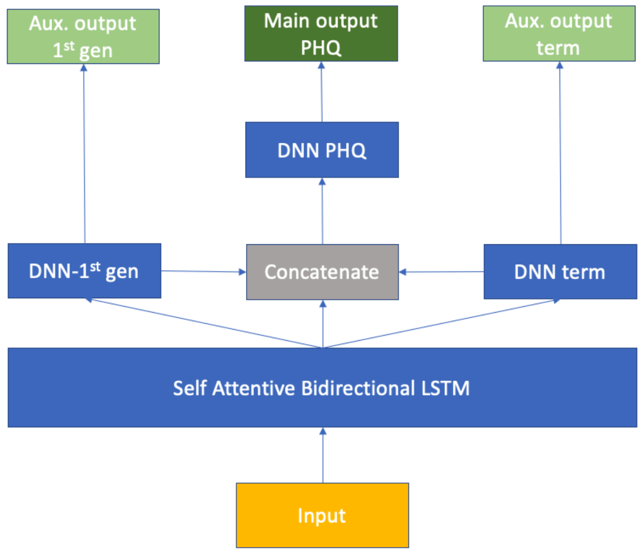


Fig. 4. Adding two auxiliary output in a neural network architecture where the losses are provided by the labels of whether the student is a first-generation student and the academic/college life progress.

Table 10. The details of DNN-1st-gen, DNN-term and DNN-PHQ in the architecture.

layer	hidden units	activation function	dropout rate
DNN-1st-gen			
output from self-attentive bi-LSTM	512		
hidden fully connected	512	ReLU	0.5
hidden fully connected	128	ReLU	
auxiliary output	1	sigmoid	
DNN-term			
output from self-attentive bi-LSTM	512		
hidden fully connected	512	ReLU	0.5
hidden fully connected	128	ReLU	
auxiliary output	1	sigmoid	
DNN-PHQ			
output from concatenate layer	768		
hidden fully connected	1024	ReLU	0.5
hidden fully connected	512	ReLU	0.5
hidden fully connected	256	ReLU	0.5
hidden fully connected	128	ReLU	
main output	1	sigmoid	

To enable the model to pay attention to differences between the first-generations and non-first-generations, we add two auxiliary outputs in our neural network architecture (Figure 4). We have observed that (1) students' behaviors change in different terms, possibly due to mixed factors, e.g., academic demands and seasonal effects (Section 3.2), (2) first-generation students have different behaviors compared to the non-first-generations (Section

5.2), and (3) the association between the behavior and mental health may be different as the year progresses (Section 5.3) . Therefore, we use two auxiliary outputs, where the losses are provided by the labels of whether the student is a first-generation student and their academic progress (i.e., which term the student is in). Using these auxiliary outputs, we attempt to contextualize mental well-being. We create two independent, fully connected sub-neural networks (DNN-1st-gen and DNN-term as shown in Figure 4) to take the output of the self-attentive bidirectional LSTM and generate auxiliary outputs. The details of the DNN-1st-gen and DNN-term, in terms of the hidden layers and number of hidden units, activation functions, dropout rate, and sizes of input and output, are shown in Table 10. The output of the self-attentive bidirectional LSTM is the input of DNN-1st-gen and DNN-term. In both DNN-1st-gen and DNN-term, there are two fully connected hidden layers with Rectified Linear Unit activation functions (ReLU). The hidden units of the last hidden layer (before the auxiliary output) of DNN-1st-gen and DNN-term networks are concatenated with the original output from self-attentive bidirectional LSTM and fed into a DNN-PHQ network (see Table 10 for details), which generates the main the depression output (i.e., PHQ-4). In this way, the network is forced to learn from the behavioral sensor data with the emphasis on the specific term as well as the student group (i.e., whether the data under consideration is associated with a first-generation student).

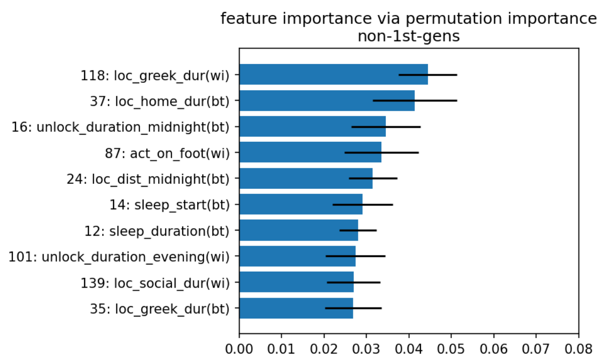
Next, we optimize the weights assigned to the main loss and auxiliary losses to improve the network's predictive performance. The loss function for each of the three outputs is binary cross-entropy. Note that the fall term is skipped due to incomplete data from the rolling enrollment process, and thus the term has only two labels. Starting from (1, 0, 0), which means the network is only trained from the main loss (i.e., PHQ-4) and does not back-propagate the loss of the auxiliary output, we gradually reduce the weight of the main output and increase the weights of the auxiliary ones. As we do not have prior knowledge of which of the two auxiliary outputs is more important, we test three different settings of the weights while keeping the main loss fixed (as noted in the same color in Table 11): (1) weight(DNN-1st-gen) > weight(DNN-term), (2) weight(DNN-1st-gen) = weight(DNN-term), and (3) weight(DNN-1st-gen) < weight(DNN-term). To prevent overfitting during training, we stratified D1 data (leave-subjects) into a training (60%) and validation (20%) set. We monitor the loss on the validation set and early-stop the training process if the loss does not decrease after more than ten iterations. Following this, the trained model is evaluated on D2 data. Table 11 shows the performance of models with different settings of weights. The results show that the F1-score firstly drops when very little auxiliary loss is added to the network. This might be because the additional auxiliary loss throws off the learning process, as it can not provide enough information because of the small weight. As we increase the weight of the auxiliary loss, the performance increases and reaches the best performance for the following settings: the weight(DNN-PHQ) = 0.7, weight(DNN-1st-gen) = 0.15, weight(DNN-term) = 0.15. The results show that with a similar weight(DNN-PHQ), the network performs best when weight(DNN-1st-gen) is approximately equal to weight(DNN-term), as shown by the dark orange, dark green and dark yellow rows in the table. Finally, the performance drops if the weight of our main output, i.e., when the weight(DNN-PHQ) is too small.

In summary, by having the learning model pay attention to whether students are first-generation or not and the academic term, the neural network improves the F1-score by 0.07, increasing performance from 0.63 to 0.70. Importantly, the proposed architecture overcomes the drawbacks of generic baseline models, which perform poorly for first-generation students. In particular, our model improves the F1-score among the first-generation students by 22%, increasing it from 0.58 to 0.71. Our approach outperforms existing models, which are biased toward the majority – non-first-generation students – while underperforming the minority first-generation students. Our approach mitigates bias and improves performance simultaneously.

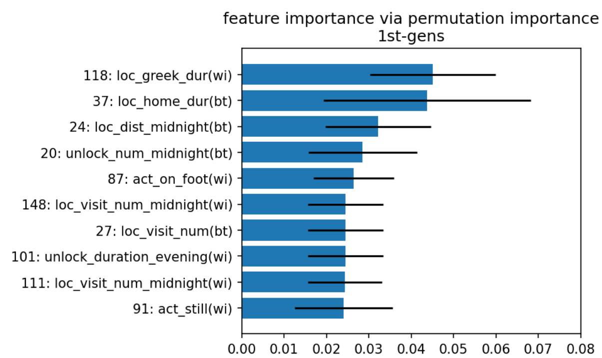
5.4.4 Interpretation. Following modeling, the natural next step is to explain better which features are essential when predicting mental health. We use the permutation feature importance algorithm [28], a model-agnostic

Table 11. Predictive performance of depression (PHQ-4) when different weights are assigned to the main loss and the auxiliary losses. Metrics are weighted to account for label imbalance.

Deep Learning weight (PHQ, 1st-gen, term)	precision	recall	F1-score	F1-score (among 1st-gen)
(1, 0, 0) – baseline	0.62	0.63	0.63	0.58
(0.95, 0.05, 0)	0.61	0.63	0.61	0.42
(0.95, 0.05, 0.05)	0.57	0.59	0.58	0.42
(0.8, 0.15, 0.05)	0.62	0.63	0.62	0.8
(0.8, 0.1, 0.1)	0.65	0.66	0.65	0.63
(0.8, 0.05, 0.15)	0.61	0.64	0.6	0.57
(0.7, 0.2, 0.1)	0.62	0.62	0.62	0.71
(0.7, 0.15, 0.15)	0.69	0.7	0.7	0.71
(0.7, 0.1, 0.2)	0.63	0.63	0.63	0.52
(0.6, 0.25, 0.15)	0.62	0.62	0.62	0.52
(0.6, 0.2, 0.2)	0.64	0.66	0.64	0.58
(0.6, 0.15, 0.25)	0.63	0.63	0.63	0.63



(a) Feature importance in predicting depression (PHQ-4) of non-first-generation students



(b) Feature importance in predicting depression (PHQ-4) of first-generation students

Fig. 5. Top 10 most important features from deep learning model discovered using permutation feature importance algorithm. The x-axis displays the decrease in the F1-score when a feature is permuted. The y-axis displays the name and index of each of the 166 features.

interpretation approach that abstracts the explanations from the deep learning model [68] in which the importance of a feature is measured by the increase in the model's prediction error after permuting the feature. The permutation is repeated 10 times for each feature. Figure 5 shows the top 10 most important features. The x-axis displays the importance score, measured by the decrease of the F1 score if a feature is permuted. The y-axis shows the feature name and index among the 166 features, with "wi" indicating a within-person standardized feature and "bt" indicating a between-person standardized feature (see Section 5.4.1 for standardizing methods). The error bars show the standard deviation of the scores across the 10 permutations for each feature. We examine the best model as discussed in Table 10 and 11. We find that time spent at various Greek houses on campus and the time spent at their own dorm are the two most important determinants in predicting mental health among non-first-generation and first-generation students, respectively. Additionally, both student groups share the common top-10 characteristics, such as activities on foot (walking/running), distance traveled around midnight, and phone use at night. Certain features are important for non-first-generation students, such as sleep onset time and sleep duration, which are both standardized among individuals. Sleep features tend to be less critical for first-generation students in comparison to non-first-generation students. The number of distinct locations visited both during the day and at night is also important. Finally, results from model-agnostic interpretation indicate the importance of features in predicting depression but not the direction; that is if the interpretation is a positive or negative association – for example, time spent at Greek Houses or time spent at dorms.

6 DISCUSSION

In this section, we discuss the main contribution of our work, the benefit of our deep learning approach, the implication of our results on better supporting first-generation students on college campuses, and finally, the limitations of the First-Gen study.

6.1 Contribution of This Work

The First-Gen study aims to paint a digital picture from phone sensor data of the first year at Dartmouth College for a small cohort of first-generation students as they rise from high school. We use mobile sensing to capture objective behavioral data over the entire first year. This research makes several contributions, including systems design of a low burden mobile sensing platform capable of high compliance and operating over an entire year; offering new research insights into risk factors and the mental health of first-generation students; and the development of a novel deep learning architecture to accurately predict mental health with a focus on first-generation students.

From the perspective of systems development, we improve the design of the StudentLife app by creatively integrating VoIP push notifications into the iOS sensing app, ensuring high compliance and low battery cost. Our First-Gen dashboard monitors all students' data compliance, allowing the research assistants "data sitting" the study to reach out to students instantly when they observe missing data. The two pilot studies to collect user input are important to ensure the robustness of the data collection phase of the study. As a result, we observe significantly higher retention and data quality than in previous long-term mobile sensing research. We believe that other researchers in this field can benefit from our experience and insights. From the perspective of student well-being, we investigate the association between various pre-college risk factors, on-campus behavior and mental health across different terms and breaks during the first year. Although previous studies explored many risk factors associated with first-generation students, they are purely based on surveys and periodic self-reports. We collect 24/7 data across the year, offering unprecedented data and insights in a passive, continuous manner. Prior first-gen studies do not include on-campus longitudinal sensing, leading them to potentially overemphasize pre-college risk factors while underestimating the adjustments students make as college life progresses. For example, we find that from the second term onward, there are no significant differences in the depression (PHQ-4) scores between first-generation and non-first-generation students. Furthermore, among the factors that first-generation students are at more risk for, according to literature, we found that only lifestyle and sociability continue affecting mental health. Our analysis suggests that rather than simply considering the initial risk factors to estimate mental health, we also need to closely note changes in lifestyle and sociability detected via mobile sensing. Our work goes beyond simply presenting depression prediction performance results and attempts to provide deeper insights through interpretation.

In terms of methodology, we propose a novel deep learning architecture to better predict the mental health of students, particularly first-generation students. Traditional machine learning approaches require us to combine group knowledge with mobile sensing features. These traditional methods do well on average, but fail to offer good performance for the minority group. The deep model presented in this paper includes group and term information as auxiliary outputs, forcing the model to pay attention to the sensing data of first-generation students. We show that by adjusting the weights associated with primary loss and auxiliary losses, we can boost the predictive performance of the group of interest (i.e., first-gens) while maintaining the performance of the broader population. We hypothesize that researchers can replace the auxiliary outputs with other factors of interest, allowing this methodology to be potentially extended to other problems.

6.2 The Benefit of Our Deep Learning Approach

There is a debate about why we need deep models when other traditional machine learning models are more interpretable and less of a black box. Although deep learning by itself does not provide interpretable features and insights, we can employ other tools such as model-agnostic interpretation approaches, such as those described in Section 5.4.4, to tease apart essential insights from the machine learning model. We find model-agnostic interpretation beneficial for longitudinal time-series behavioral data from phones. Traditional machine learning algorithms struggle with time-series mobile sensing data, where researchers engineer hand-crafted features or aggregates – statistical features on each time-series data (e.g., average, percentile, maximum, minimum, etc.) – resulting in loss of information and an explosion of highly correlated features. There is no easy way to assign weights to different days in the time series because we do not know which day is more relevant to predicting a specific outcome. Deep learning algorithms, such as LSTM, are adept at handling such time series challenges.

The use of self-attention enables learning models to learn the weights of days from the sensor data directly, avoiding the need for prior information. Figure 6 shows an attention vector indicating the weights of different days during a term to predict depression (PHQ-4) for a specific student. In this case, the behavior before and after the midterm period is more important, maybe because other students exhibit similar changes in behavior during midterms, making data collected during the midterm less valuable in distinguishing overall good or bad mental health during the term. Non-deep learning models implicitly treat the whole term evenly. The use of self-attention in this regard improves model generalizability because self-attention can learn contextual cues from a specific population under study and their experiences. For example, consider the case of different schools having different schedules. A deep model can be trained and customized to the context of a specific school from sensing data without being bound by certain statistical aggregation of the data.

Finally, deep learning allows us to treat the information of interest as auxiliary output rather than input. Such an approach provides flexibility, as the model can be modified to perform well for a separate task as well.

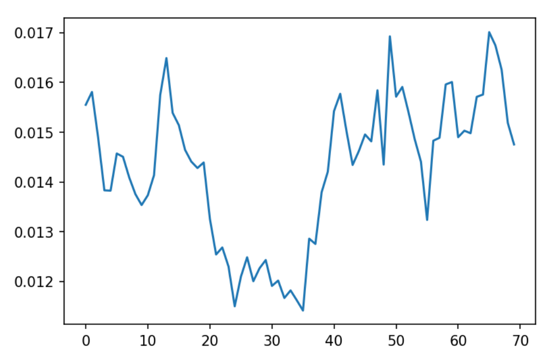


Fig. 6. The attention vector indicates the weights of different days during a term in predicting depression (PHQ-4) from sensing data from a person.

6.3 College Support for First-generation Students

Today there is a data and a knowledge gap at colleges supporting at-risk students such as first-gens. Professors have little or no visibility of struggling students outside of academics. Student deans get to know students more personally and may have better insights into how students are coping. Wellness programs inside the administrations set the policy and adopt programs to advance the health and safety of students. Health centers on campuses, both on the clinical and counseling side, are on the front line when dealing with the ebb and flow of mental health across the school year. Institutional statistics collected each year indicate that the trajectory of mental health across the country is heading in the wrong direction. We observe a disconnect between all the stakeholders in that none has a complete view of the dynamics of students' mental health. We believe the missing link or the glue involves characterizing mental health using mobile sensing.

Compared to other students, first-generation students arrive at college less prepared and with fewer resources and coping skills. As a result, first-gens struggle with the social, financial and academic demands they confront. For this reason, many colleges have active first-generation programs to help students navigate college. Dartmouth College, for example, has a month-long summer orientation program between high school and college for admitted first-generation students. During orientation, we teach mini-classes and offer skill-building workshops, self-assessment tools, and goal-setting sessions. The goal of the orientation is to Bootcamp their first year, where students learn about the pitfalls, hear about coping skills, talk to faculty and begin to forge a strong peer network that will help carry them through the college years.

Beyond college programs for first-gens, researchers are making progress in designing new mental health technologies to support college students. However, we are in the early stages of this development. A recent study presents a series of design activities conducted with college students and emphasizes the social ecosystems and social support networks in a college student's life [45]. We believe StudentLife and the First-Gen study could potentially open up new insights and ideas for moving things forward. For example, new forms of intervention may emerge to help keep students healthy and on track academically. Combining behavioral sensing from phones and wearables with new intervention systems is an open and important area of research.

In our study, we train a model using mobile sensing data and self-reported EMAs. After training the network using the available labels and validating the model's generalizability, we might eliminate the need for self-reports/EMAs, as the model should be capable of estimating the mental health of new individuals not in the training data. The results discussed in this paper provide objective insights and, as a next step, could provide possible suggestions to first-gens associated with alleviating poor behavioral patterns by promoting behaviors that lead to better mental health. With such an AI-based sensing and intervention system, risks would be detected early. More importantly, we envision personalized and accurate interventions that provide suggestions based on behavioral sensing data.

Although our work is a proof-of-concept, we believe that the method presented in the paper can be scaled to a larger population with ease. First, it could target individuals or subgroups (say, first-gens) whose risk of developing depression is higher based on the risk factors. Second, through early detection, we can provide help and guidance to them to ensure that they have a better college experience. And finally, with the understanding of what works and what does not in helping students, we will be able to help reduce the future recurrence of poor behaviors that negatively impact mental health. Future mental health sensing and intervention platforms will offer alerts, guidance, and tips anytime and anywhere, which may be ideal for students who have trouble with in-person appointments. The machine learning model itself can be run offline or on-device to help protect the privacy of students. The modern phones are equipped with neural processing units and high-efficiency edge computing hardware that makes the model training and prediction fast and scalable. We can also perform the preprocessing either locally or leverage edge or cloud computing. The future of mental health sensing will combine phones and wearables to offer higher performance prediction and personalized sensing and intervention.

Such a combined platform would integrate physiological signals, behavioral sensing, and contextual insights (i.e., around the student).

6.4 Limitations

Although our study offers many interesting insights into the behavior of first-gen students, it also comes with some limitations that should be noted. First, our study is conducted at a single university. Therefore, our findings may not be consistent with first-generation students in another university. We do not make any claims of generalization of our results. We can say that we found important differences between first-gens and non-first-generation students at Dartmouth College. However, we believe that our methodology has the potential to be generalized, as we discussed in Section 6.2. Only through careful replication and reproducibility will we be able to understand best if our results offer general insights.

Second, we enlisted as many students as possible as a university research team, given staffing and funding limitations. However, we must acknowledge that the small size of the first-generation student population may result in false discovery and overfitting. To limit the likelihood of drawing an inaccurate result, we employ rigorous statistical methods (e.g., splitting the data into two random, non-overlapping subsets, FDR and leave-subjects-out test group) to demonstrate the potential of such an approach. Even so, the result and approach must be retested with a bigger sample to determine its generalizability.

Another limitation is the reliability of the iOS/Android activity recognition APIs used to infer physical activities. Since different companies manufacture Android devices, mobile phones may have a vendor-specific hardware sensor and software implementation that might lead to some errors in activity recognition. Additionally, we recognize that the recognition algorithm may evolve across iOS/Android versions. All of these factors would result in inconsistency in activity recognition. Although Apple and Google assert that they ensure accuracy through training on vast amounts of data, it is quite difficult for us to quantify any inherent error in these measurements. This may continue to be a restriction for all studies based on smartphone and wearable sensing, as we have yet to find literature that supports this inaccuracy and resolves any problems associated with it.

Furthermore, even though smartphone sensing can capture various behaviors, it is still limited in capturing very fine-grained physical activities. For instance, students may set their phones down while running on a treadmill, in which case smartphone sensors would be unable to recall this physical activity (on-campus location can serve as a better proxy). In general, one can expect the physical activities (walking / running / cycling / in-vehicle / sedentary) to be more accurate outdoors when participants carry their phones on their bodies and less accurate indoors when participants are more likely to place their phones on a table. In other studies [52, 58, 72], researchers examine alternative sensing modalities (e.g., wearables, in-house sensors) for more precise activity identification to circumvent the limitation of smartphone sensing. However, we did not pursue this course of action in our study because we are focused on better understanding how effectively a low-cost tool, such as StudentLife, predicts the mental health of at-risk students.

7 CONCLUSION

In this paper, we used the StudentLife mobile sensing platform to track the behavior of N=180 first-year students at Dartmouth college for an entire year. We assessed the risk factors of all students focusing specifically on first-generation students who are more at risk than other groups. We considered various factors associated with first-year students' risk, including socioeconomic status, lifestyle, sociability and support network. We discussed the First-Gen year-long study to investigate how first-generation students' behaviors are associated with their depression and anxiety scores. We observed differences in sensed behaviors of students across each term and academic break. We also reported on behavioral differences between first-generation and non-first-generation students. We designed a novel deep neural network architecture that can learn more informative insights from the behaviors of first-generation students, providing more accurate mental health predictions than generic models. We are currently following the first-generation students in this study across their remaining college years and hope to report our findings in the future after they graduate.

ACKNOWLEDGMENTS

This work is supported by National Institute of Mental Health (NIMH), grant number 5R01MH059282. Each year Dartmouth College admits approximately 1000 undergraduates, where 10 to 12 % of the admitted class are first-generation students. The authors of this paper are grateful to Jay Davis, who inspires and directs Dartmouth College's first generation program. We are also thankful to all the amazing first-generation students at Dartmouth College who inspired our work. In particular, this paper would not have been possible without the efforts of Lessley Hernandez and Vlado Vojdanovski, who are first gen students and co-authors of this paper. Finally, Andrew Campbell is a proud first-generation student circa the 1980s.

REFERENCES

[1] Saeed Abdullah, Mark Matthews, Ellen Frank, Gavin Doherty, Geri Gay, and Tanzeem Choudhury. 2016. Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association* 23, 3 (2016), 538–543.

[2] Daniel A Adler, Dror Ben-Zeev, Vincent WS Tseng, John M Kane, Rachel Brian, Andrew T Campbell, Marta Hauser, Emily A Scherer, and Tanzeem Choudhury. 2020. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. *JMIR mHealth and uHealth* 8, 8 (2020), e19962.

[3] Apple 2021. Extending Your App's Background Execution Time. https://developer.apple.com/documentation/uikit/app_and_environment/scenes/preparing_your_ui_to_run_in_the_background/extending_your_app_s_background_execution_time.

[4] AWARE framework 2013. Aware: Open-source Context Instrumentation Framework For Everyone. http://www.awareframework.com/.

[5] Yoav Benjamini, Abba M Krieger, and Daniel Yekutieli. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 3 (2006), 491–507.

[6] Janet Mancini Billson and Margaret Brooks Terry. 1982. In Search of the Silken Purse: Factors in Attrition among First-Generation Students. Revised. (1982).

[7] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex (Sandy) Pentland. 2014. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 477–486. https://doi.org/10.1145/2647868.2654933

[8] Mehdi Boukhechba, Alexander R. Daros, Karl Fua, Philip I. Chow, Bethany A. Teachman, and Laura E. Barnes. 2018. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health* (July 2018). https://doi.org/10.1016/j.smhl.2018.07.005

[9] Mehdi Boukhechba, Yu Huang, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. 2017. Monitoring social anxiety from mobility and communication patterns. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 749–753.

[10] Pierre Bourdieu, John G Richardson, et al. 1986. Handbook of Theory and Research for the Sociology of Education. *The forms of capital* (1986), 241–258.

[11] Noli Brazil and Matthew Andersson. 2018. Mental Well-Being and Changes in Peer Ability From High School to College. *Youth & Society* (March 2018), 0044118X1876452. https://doi.org/10.1177/0044118X18764526

[12] Ian Brissette, Michael F Scheier, and Charles S Carver. 2002. The role of optimism in social network development, coping, and psychological adjustment during a life transition. *Journal of personality and social psychology* 82, 1 (2002), 102.

[13] Khanh Van T Bui. 2002. First-generation college students at a four-year university: Background characteristics, reasons for pursuing higher education, and first-year experiences. *College Student Journal* 36, 1 (2002), 3–12.

[14] Elizabeth A Canning, Jennifer LaCosse, Kathryn M Kroeper, and Mary C Murphy. 2020. Feeling like an imposter: the effect of perceived classroom competition on the daily psychological experiences of first-generation college students. *Social Psychological and Personality Science* 11, 5 (2020), 647–657.

[15] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 1293–1304.

[16] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 145–152.

[17] Colleen S Conley, Alexandra C Kirsch, Daniel A Dickson, and Fred B Bryant. 2014. Negotiating the transition to college: Developmental trajectories and gender differences in psychological functioning, cognitive-affective strategies, and social well-being. *Emerging Adulthood* 2, 3 (2014), 195–210.

[18] Catherine M Coveney. 2014. Managing sleep and wakefulness in a 24-hour world. *Sociology of health & illness* 36, 1 (2014), 123–136.

[19] Robert Crosnoe and Chandra Muller. 2014. Family socioeconomic status, peers, and the path to college. *Social problems* 61, 4 (2014), 602–624.

[20] Kadir Demirci, Mehmet Akgönül, and Abdullah Akpinar. 2015. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. *Journal of behavioral addictions* 4, 2 (2015), 85–92.

[21] Jessica M Dennis, Jean S Phinney, and Lizette Ivy Chuateco. 2005. The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of college student development* 46, 3 (2005), 223–236.

[22] Kevin Eagan, Ellen Bara Stolzenberg, Joseph J Ramirez, Melissa C Aragon, Maria Ramirez Suchard, and Sylvia Hurtado. 2014. The American freshman: National norms fall 2014. *Los Angeles: Higher Education Research Institute, UCLA* (2014).

[23] Fifth Edition et al. 2013. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc* 21 (2013), 591–643.

[24] Daniel Eisenberg, Ezra Golberstein, and Justin B Hunt. 2009. Mental health and academic success in college. *The BE Journal of Economic Analysis & Policy* 9, 1 (2009).

[25] Walid El Ansari and Christiane Stock. 2010. Is the health and wellbeing of university students associated with their academic performance? Cross sectional findings from the United Kingdom. *International journal of environmental research and public health* 7, 2 (2010), 509–527.

[26] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.

[27] Farhan, Yue, Morillo, Ware, Lu, Bi, Kamath, Russell, Bamis, and Wang. 2016. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In *2016 IEEE Wireless Health (WH)*. IEEE, Bethesda, MD, USA, 1–8. https://doi.org/10.1109/WH.2016.7764553

[28] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. 2019. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* 20, 177 (2019), 1–81.

[29] Kim Fromme, William R. Corbin, and Marc I. Kruse. 2008. Behavioral Risks during the Transition from High School to College. *Developmental psychology* 44, 5 (Sept. 2008), 1497–1504. https://doi.org/10.1037/a0012614

[30] Google Activity Recognition Api. 2019. Google Activity Recognition Api. https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionClient.

[31] Gabriella M Harari, Samuel D Gosling, Rui Wang, Fanglin Chen, Zhenyu Chen, and Andrew T Campbell. 2017. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior* 67 (2017), 129–138.

[32] Gabriella M Harari, Nicholas D Lane, Rui Wang, Benjamin S Crosier, Andrew T Campbell, and Samuel D Gosling. 2016. Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* 11, 6 (2016), 838–854.

[33] Douglas N Harris. 2010. How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers college record* 112, 4 (2010), 1163–1197.

[34] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[35] Cassandra Holinka. 2015. Stress, emotional intelligence, and life satisfaction in college students. *College Student Journal* 49, 2 (2015), 300–311.

[36] Joann Horton. 2015. Identifying at-risk factors that affect college student success. *International Journal of Process Education* 7, 1 (2015), 83–101.

[37] Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A. Teachman, and Laura E. Barnes. 2016. Assessing Social Anxiety Using Gps Trajectories and Point-of-interest Data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 898–903. https://doi.org/10.1145/2971648.2971761

[38] Jeremy F Huckins, Alex W DaSilva, Rui Wang, Weichen Wang, Elin L Hedlund, Eilis I Murphy, Richard B Lopez, Courtney Rogers, Paul E Holtzheimer, William M Kelley, et al. 2019. Fusing mobile phone sensing and brain imaging to assess depression in college students. *Frontiers in neuroscience* 13 (2019), 248.

[39] iOS Core Motion. 2019. iOS Core Motion. https://developer.apple.com/documentation/coremotion.

[40] Terry T Ishitani. 2003. A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. *Research in higher education* 44, 4 (2003), 433–449.

[41] Sharon Rae Jenkins, Aimee Belanger, Melissa Londoño Connally, Adriel Boals, and Kelly M Durón. 2013. First-generation undergraduate students' social support, depression, and life satisfaction. *Journal of College Counseling* 16, 2 (2013), 129–142.

[42] Laura E. Knouse, Greg Feldman, and Emily J. Blevins. 2014. Executive functioning difficulties as predictors of academic performance: Examining the role of grade goals. *Learning and Individual Differences* 36 (Dec. 2014), 19–26. https://doi.org/10.1016/j.lindif.2014.07.001

[43] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[44] Kurt Kroenke, Robert L Spitzer, Janet BW Williams, and Bernd Löwe. 2009. An ultra-brief screening scale for anxiety and depression: the PHQ–4. *Psychosomatics* 50, 6 (2009), 613–621.

[45] Emily G Lattie, Rachel Kornfield, Kathryn E Ringland, Renwen Zhang, Nathan Winquist, and Madhu Reddy. 2020. Designing Mental Health Technologies that Support the Social Ecosystem of College Students. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[46] Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. 707–710.

[47] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. (2013), 13.

[48] Hubert W Lilliefors. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association* 62, 318 (1967), 399–402.

[49] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. *arXiv preprint arXiv:1703.03130* (2017).

[50] Mary J Lindstrom and Douglas M Bates. 1988. Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Statist. Assoc.* 83, 404 (1988), 1014–1022.

[51] Herbert Warren Marsh. 2006. Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology. British Psychological Society London.

[52] Stephen M Mattingly, Julie M Gregg, Pino Audia, Ayse Elvan Bayraktaroglu, Andrew T Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Sidney K D'Mello, Anind K Dey, et al. 2019. The Tesserae project: Large-scale, longitudinal, in situ, multimodal sensing of information workers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[53] Graziella Pagliarulo McCarron and Karen Kurotsuchi Inkelas. 2006. The gap between educational aspirations and attainment for first-generation college students and the role of parental involvement. *Journal of College Student Development* 47, 5 (2006), 534–549.

[54] Deanna LH McFadden. 2016. Health and academic success: A look at the challenges of first-generation community college students. *Journal of the American Association of Nurse Practitioners* 28, 4 (2016), 227–232.

[55] Abhinav Mehrotra and Mirco Musolesi. 2017. Designing Effective Movement Digital Biomarkers for Unobtrusive Emotional State Mobile Monitoring. (2017).

[56] Sanjay S Mehta, John J Newbold, and Matthew A O'Rourke. 2011. Why do first-generation students fail? *College Student Journal* 45, 1 (2011), 20–36.

[57] Microsoft 2020. AppCenter platform. https://appcenter.ms/.

[58] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K Dey, et al. 2019. Differentiating higher and lower job performers in the workplace using mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–24.

[59] Sean P Mullen, John F Adamek, Madhura Phansikar, Brent Roberts, and Christopher Larrison. 2020. A Path Analysis of the Role of First-Generation Status and Engagement in Social Interaction, Physical Activity, and Therapy in Satisfaction with Life among College Students. (2020).

[60] Subigya Nepal, Gonzalo J. Martinez, Shayan Mirjafari, Stephen Mattingly, Vedant Das Swain, Aaron Striegel, Pino G. Audia, and Andrew T. Campbell. 2021. Assessing the Impact of Commuting on Workplace Performance Using Mobile Sensing. *IEEE Pervasive Computing* 20, 4 (Oct. 2021), 52–60. https://doi.org/10.1109/mprv.2021.3112399

[61] Subigya Nepal, Shayan Mirjafari, Gonzalo J. Martinez, Pino Audia, Aaron Striegel, and Andrew T. Campbell. 2020. Detecting Job Promotion in Information Workers Using Mobile Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 113 (sep 2020), 28 pages. https://doi.org/10.1145/3414118

[62] Subigya Nepal, Weichen Wang, Vlado Vojdanovski, Jeremy F Huckins, Alex daSilva, Meghan Meyer, and Andrew Campbell. 2022. COVID Student Study: A Year in the Life of College Students during the COVID-19 Pandemic Through the Lens of Mobile Phone Sensing. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 42, 19 pages. https://doi.org/10.1145/3491102.3502043

[63] Mikio Obuchi, Jeremy F Huckins, Weichen Wang, Alex daSilva, Courtney Rogers, Eilis Murphy, Elin Hedlund, Paul Holtzheimer, Shayan Mirjafari, and Andrew Campbell. 2020. Predicting Brain Functional Connectivity Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22.

[64] Venet Osmani, Alban Maxhuni, Agnes Grünerbl, Paul Lukowicz, Christian Haring, and Oscar Mayora. 2013. Monitoring activity of patients with bipolar disorder using smart phones. In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*. 85–92.

[65] Skyler Place, Danielle Blanch-Hartigan, Channah Rubin, Cristina Gorrostieta, Caroline Mead, John Kane, Brian P Marx, Joshua Feast, Thilo Deckersbach, Andrew Nierenberg, et al. 2017. Behavioral indicators on a mobile sensing platform predict clinically validated

psychiatric symptoms of mood and anxiety disorders. *Journal of medical Internet research* 19, 3 (2017), e75.

[66] Lucila Ramos-Sánchez and Laura Nichols. 2007. Self-efficacy of first-generation and non-first-generation college students: The relationship with academic performance and college adjustment. *Journal of college counseling* 10, 1 (2007), 6–18.

[67] Gerald M Reid, Melissa K Holt, Chelsey E Bowman, Dorothy L Espelage, and Jennifer Greif Green. 2016. Perceived social support and mental health among first-year college students with histories of bullying victimization. *Journal of Child and Family Studies* 25, 11 (2016), 3331–3341.

[68] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386* (2016).

[69] Sylvia Ruiz, Jessica Sharkness, Kimberly Kelly, Linda DeAngelo, and John Pryor. 2010. Findings from the 2009 administration of the Your First College Year (YFCY): National aggregates. *Los Angeles: Higher Education Research Institute at the University of California Los Angeles* (2010).

[70] Sohrab Saeb, Emily G Lattie, Stephen M Schueller, Konrad P Kording, and David C Mohr. 2016. The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* 4 (2016), e2537.

[71] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research* 17, 7 (2015).

[72] Koustuv Saha, Manikanta D Reddy, Vedant das Swain, Julie M Gregg, Ted Grover, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. Imputing missing social media data stream in multisensor studies of human behavior. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 178–184.

[73] A. Sano, A. J. Phillips, A. Z. Yu, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman, and R. W. Picard. 2015. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 1–6. https://doi.org/10.1109/BSN.2015.7299420

[74] Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45, 11 (1997), 2673–2681.

[75] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th International Conference on World Wide Web*. 103–112.

[76] Jessica Sharkness and Linda DeAngelo. 2011. Measuring student involvement: A comparison of classical test theory and item response theory in the construction of scales from student surveys. *Research in Higher Education* 52, 5 (2011), 480–507.

[77] Giovanni Sogari, Catalina Velez-Argumedo, Miguel I Gómez, and Cristina Mora. 2018. College students and eating habits: A study using an ecological model for healthy behavior. *Nutrients* 10, 12 (2018), 1823.

[78] Patrick T Terenzini, Leonard Springer, Patricia M Yaeger, Ernest T Pascarella, and Amaury Nora. 1996. First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher education* 37, 1 (1996), 1–22.

[79] Jordan Thibodeaux, Aaron Deutsch, Anastasia Kitsantas, and Adam Winsler. 2017. First-Year College Students' Time Use: Relations With Self-Regulation and GPA. *Journal of Advanced Academics* 28, 1 (Feb. 2017), 5–27. https://doi.org/10.1177/1932202X16676860

[80] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288.

[81] Vincent W-S Tseng, Akane Sano, Dror Ben-Zeev, Rachel Brian, Andrew T Campbell, Marta Hauser, John M Kane, Emily A Scherer, Rui Wang, Weichen Wang, et al. 2020. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. *Scientific reports* 10, 1 (2020), 1–17.

[82] John Von Neumann, RH Kent, HR Bellinson, and BI t Hart. 1941. The mean square successive difference. *The Annals of Mathematical Statistics* 12, 2 (1941), 153–162.

[83] MaryBeth Walpole. 2003. Socioeconomic status and college: How SES affects college experiences and outcomes. *The review of higher education* 27, 1 (2003), 45–73.

[84] 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*. ACM, 3–14.

[85] Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T. Campbell. 2015. SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 295–306. https://doi.org/10.1145/2750858.2804251 event-place: Osaka, Japan.

[86] Rui Wang, Weichen Wang, Min SH Aung, Dror Ben-Zeev, Rachel Brian, Andrew T Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Emily A Scherer, et al. 2017. Predicting Symptom Trajectories of Schizophrenia using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 110.

[87] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proceedings of the ACM on Interactive,*

Mobile, Wearable and Ubiquitous Technologies 2, 1 (March 2018), 1–26. https://doi.org/10.1145/3191775

[88] Weichen Wang, Gabriella M Harari, Rui Wang, Sandrine R Müller, Shayan Mirjafari, Kizito Masaba, and Andrew T Campbell. 2018. Sensing Behavioral Change over Time: Using Within-Person Variability Features from Mobile Sensing to Predict Personality Traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 141.

[89] Weichen Wang, Shayan Mirjafari, Gabriella Harari, Dror Ben-Zeev, Rachel Brian, Tanzeem Choudhury, Marta Hauser, John Kane, Kizito Masaba, Subigya Nepal, et al. 2020. Social sensing: assessing social functioning of patients living with schizophrenia using mobile phone sensing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[90] Weichen Wang, Jialing Wu, Subigya Kumar Nepal, Alex daSilva, Elin Hedlund, Eilis Murphy, Courtney Rogers, and Jeremy F. Huckins. 2021. On the Transition of Social Interaction from In-Person to Online: Predicting Changes in Social Media Usage of College Students during the COVID-19 Pandemic Based on Pre-COVID-19 On-Campus Colocation. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 425–434. https://doi.org/10.1145/3462244.3479888

[91] Karl R White. 1982. The relation between socioeconomic status and academic achievement. *Psychological bulletin* 91, 3 (1982), 461.

[92] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K Villalba, Janine M Dutcher, Michael J Tumminia, Tim Althoff, Sheldon Cohen, Kasey G Creswell, J David Creswell, et al. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–33.

[93] Dollean C York-Anderson and Sharon L Bowman. 1991. Assessing the college knowledge of first-generation and second-generation college students. *Journal of College Student Development* (1991).

[94] Han Yu and Akane Sano. 2020. Passive Sensor Data Based Future Mood, Health, and Stress Prediction: User Adaptation Using Deep Learning. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 5884–5887.

[95] Xiao Zhang, Wenzhong Li, Xu Chen, and Sanglu Lu. 2018. MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (Jan. 2018), 1–30. https://doi.org/10.1145/3161414