Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience and Behavior of College Students during the Pandemic

SUBIGYA NEPAL, Dartmouth College, USA
WENJUN LIU, Dartmouth College, USA
ARVIND PILLAI, Dartmouth College, USA
WEICHEN WANG, Dartmouth College, USA
VLADO VOJDANOVSKI, Dartmouth College, USA
JEREMY F. HUCKINS, Biocogniv Inc, USA
COURTNEY ROGERS, Dartmouth College, USA
MEGHAN L. MEYER, Columbia University, USA
ANDREW T. CAMPBELL, Dartmouth College, USA

Understanding the dynamics of mental health among undergraduate students across the college years is of critical importance, particularly during a global pandemic. In our study, we track two cohorts of first-year students at Dartmouth College for four years, both on and off campus, creating the longest longitudinal mobile sensing study to date. Using passive sensor data, surveys, and interviews, we capture changing behaviors before, during, and after the COVID-19 pandemic subsides. Our findings reveal the pandemic's impact on students' mental health, gender based behavioral differences, impact of changing living conditions and evidence of persistent behavioral patterns as the pandemic subsides. We observe that while some behaviors return to normal, others remain elevated. Tracking over 200 undergraduate students from high school to graduation, our study provides invaluable insights into changing behaviors, resilience and mental health in college life. Conducting a long-term study with frequent phone OS updates poses significant challenges for mobile sensing apps, data completeness and compliance. Our results offer new insights for Human-Computer Interaction researchers, educators and administrators regarding college life pressures. We also detail the public release of the de-identified *College Experience Study* dataset used in this paper and discuss a number of open research questions that could be studied using the public dataset.

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

ACM Reference Format:
Subigya Nepal, Wenjun Liu, Arvind Pillai, Weichen Wang, Vlado Vojdanovski, Jeremy F. Huckins, Courtney Rogers, Meghan L. Meyer, and Andrew T. Campbell. 2024. Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience and Behavior of College Students during the Pandemic. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 38 (March 2024), 37 pages. https://doi.org/10.1145/3643501

Authors' addresses: Subigya Nepal, Dartmouth College, Department of Computer Science, Hanover, NH, USA, sknepal@cs.dartmouth.edu; Wenjun Liu, Dartmouth College, Department of Computer Science, Hanover, NH, USA; Arvind Pillai, Dartmouth College, Department of Computer Science, Hanover, NH, USA; Weichen Wang, Dartmouth College, Department of Computer Science, Hanover, NH, USA; Vlado Vojdanovski, Dartmouth College, Hanover, NH, USA; Jeremy F. Huckins, Biocogniv Inc, Burlington, VT, USA; Courtney Rogers, Dartmouth College, Psychological and Brain Sciences, Hanover, NH, USA; Meghan L. Meyer, Columbia University, Department of Psychology, New York, NY, USA; Andrew T. Campbell, Dartmouth College, Department of Computer Science, Hanover, NH, USA.

2474-9567/2024/3-ART38
https://doi.org/10.1145/3643501

1 INTRODUCTION

The college years represent a critical period in students' lives, marked by significant personal, social, and academic growth [6, 36, 79]. Understanding the behavioral patterns and mental health of college students during this time is essential for promoting their well-being and success. Data-driven approaches, such as mobile sensing techniques and Ecological Momentary Assessment (EMA) surveys, have emerged as valuable tools for investigating the link between daily activities and mental health [19, 23, 72]. By analyzing sensor data and survey responses, researchers can uncover trends in students' daily activities, behavioral trends, social interactions, and emotional states, shedding light on the factors that influence their mental health and overall well-being. The landscape of college life has dramatically shifted in recent years due to the COVID-19 pandemic and subsequent national lockdowns [65, 67]. These changes have presented unprecedented challenges for students, highlighting the need to better understand how they affect mental health. Longitudinal studies that examine students' behaviors and experiences throughout their college career, encompassing both the pre-pandemic years and the impact of COVID-19, can provide a deep understanding of the factors that contribute to their mental health and resilience. Such studies would offer valuable insights for supporting students during both normal and extraordinary times.

In this paper, we present the findings of a comprehensive, longitudinal study that leverages both passive sensing data and EMA surveys to explore the experiences of two cohorts of undergraduate students at Dartmouth College over a period of five years, spanning from September 2017 to June 2022. Our analysis encompasses various aspects of college life, including pre-pandemic years, the COVID-19 pandemic, and the gradual return to normalcy as the pandemic subsided. By examining behavioral patterns and mental health metrics during different periods, we aim to uncover the factors that contribute to students' well-being and resilience, and to provide valuable insights that can inform the development of effective interventions and support systems for college students. This study utilizes the StudentLife mobile application [73] used and validated in many previous studies [10, 25, 51, 54, 75] to facilitate the consistent and fine-grained collection of passive sensing data and EMA surveys from undergraduate students over an extended period. By gathering objective data at varying intervals, in addition to weekly self-reported EMA surveys, we are able to paint a detailed picture of students' daily activities, behaviors, mental health, and overall college experience. The rich dataset provided by the sensing app allows us to explore various aspects of college life, such as behavioral differences based on gender, changes in living arrangements (on-campus vs. off-campus), and the persistence of behavioral patterns after the pandemic subsides.

Our study not only examines the impact of the COVID-19 pandemic on students' mental health and daily activities but also delves into the broader college experience over four years. By analyzing pre-pandemic data, we can identify trends in students' mental health and behavior prior to COVID and across their college years. This comprehensive approach enables us to understand the resilience and adaptability of students in the face of unprecedented challenges, as well as the factors that may contribute to their mental well-being. In particular, we found evidence that events during the COVID-19 pandemic influenced students' concern levels over time, as demonstrated by the link between data evaluating concern and mental status and events. We also observed differences in behavioral patterns between the first and last academic terms of both cohorts, highlighting changes in behavior over the course of students' college years. The first cohort of $N = 104$ students started college in Fall 2017 and graduated in Spring 2021, while the second cohort of $N = 113$ students started in Fall 2018 and graduated in Spring 2022. We also identified differences in behavioral patterns in different phases (pre-COVID, during-COVID and in-person instruction which is the period when the students returned to campus after COVID subsided), suggesting the effect of the pandemic on human behavior change. In addition to extensive quantitative analysis of data collected through our mobile sensing app, we also conducted qualitative interviews with fifteen students who graduated in 2022 and 2023. These interviews provide additional insights into the impact of the COVID-19 pandemic and other factors on college students' mental health and daily activities. By integrating both quantitative and qualitative data, our study offers a comprehensive understanding of the interplay between daily activities, mental health, and resilience during the college years. This mixed-methods approach allows us to identify patterns and trends in the data while capturing the richness and diversity of students' experiences. The paper makes a number of contributions including:

(1) **Longitudinal Study and Analysis of the College Years**: Our study presents a rare longitudinal analysis of undergraduate students, offering valuable insights into their mental health and daily activities over an extended period. We visualize the sensing data and self-reports that show insightful trends, such as a spike in student stress and anxiety levels in the run-up to returning to campus from breaks. This increase might be tied to the impending return to campus and academic burdens. Our study and analysis shed further light on the challenges students faced during their college years, including the rising risk during the COVID-19 pandemic.
(2) **Factors Influencing Student Mental Health**: We delve into the influence of various factors on students' mental health and their overall college experience. We examine behavioral differences based on gender, comparing the first term to the final term, and consider changes in living arrangements. For example, we find that female students exhibit higher sedentary behavior, elevated COVID concern, higher social level, but lower self-esteem. We also identify significant behavioral and mental shifts between the first and last academic terms, highlighting differences in behavior between the initial and final semesters.
(3) **Impact of the COVID-19 Pandemic on College Life**: We explore the persistence of behavioral patterns after the COVID-19 pandemic subsides, aiming to provide insights for the development of targeted interventions and support systems. Our study reveals a noticeable negative impact of the pandemic on students' mental health, with increased levels of depression, anxiety, stress, and psychological distress. While certain behaviors, such as physical activity and time spent at specific locations, are returning to pre-pandemic levels, others, such as sleep duration, remain elevated, and social interaction is reduced. These findings emphasize the need for additional support during the transition back to normal routines.
(4) **Qualitative Insights from Undergraduates**: To deepen our understanding of the college experience during this unique period, we incorporate qualitative interviews with graduating students. These interviews offer insight into the challenges and successes students encountered while navigating their college years amid a global pandemic.
(5) **College Experience Study Dataset and Open Research Questions**: In the interest of advancing research into the mental health of college students during their college years using mobile sensing and self-reports, we discuss the public release of the de-identified College Experience Study dataset used in this paper. We also outline a series of open research questions that could be studied using the public dataset. The College Experience Study dataset is available here [26].

The *College Experience Study* is the longest mobile sensing study to date, investigating the relationship between daily activities, behaviors, mental health, and resilience during the college years. The findings have significant implications for interventions, policies, and support systems promoting student well-being and success in higher education, especially during a global pandemic. In this study, our primary focus is on examining the experiences of college students at a single institution in the US, recognizing that our findings may be uniquely applicable to this specific demographic group. As a result, we underscore the need for further research to assess the relevance of our insights across various populations and settings.

This paper presents related work, background information, and an overview of the dataset in Sections 2 and 3. It explores the collected data, including the impact of COVID-19, in Section 4. Detailed observations on differences between the first and final term, gender, changes in living arrangements, and persistent behavioral patterns during the pandemic are provided in Sections 5, 6, 7, and 8, respectively. We discuss the findings, implications and ethical considerations in Section 9. This is followed by limitations and future work in Section 10. The public release of the de-identified dataset and open research questions that could be studied with the dataset are discussed in Section 11. We conclude in Section 12 with reflections on the study and some final remarks.

2 RELATED WORK

Mobile sensing has surfaced as a prevalent tool for passively analyzing human behavior [1, 30, 33, 60]. A considerable volume of research has established connections between mobile sensing and diverse facets of personality [62, 80], mental health [2, 3, 9, 20, 50, 58, 64], and professional behavior [18, 29, 39, 47, 51–53, 57]. This encompasses associating daily activity with academic performance [64, 73, 74], identifying mental health risks [4, 21, 76], and modeling behavior for mental health assessment [32, 35, 42, 69]. Recently, data-driven methodologies for examining the relationship between daily activities and mental health have attracted significant scholarly interest [11, 15, 48]. Mobile sensing techniques, together with Ecological Momentary Assessment (EMA) surveys, have been widely employed to gather behavioral data, thereby formulating models that encapsulate mental health parameters, specifically targeting student demographics [35, 40, 41, 66, 77, 83]. The COVID-19 pandemic has underscored the potential of mobile sensing and data analytics for tracking the progression of diseases at various scales—individual, neighborhood, city, and national levels. Mobile sensing has evolved into an indispensable resource for identifying personal infection statuses, monitoring long-term health shifts, tracing community epidemics, and observing the development of viruses and their variants [78]. The passive, in-situ data gathering facilitated by mobile sensing offers the distinct advantage of enabling factual deductions in real-world settings, thus proving to be a highly effective tool for studying human behavior during the pandemic. The insights derived from this methodology considerably augment our comprehension of how individuals navigate such extraordinary situations.

The COVID-19 pandemic and ensuing lockdowns have significantly affected students' lifestyles and mental well-being [13, 63, 87]. While some students have managed to quickly adapt, others have grappled with stress, anxiety, and other mental health issues during this period [25, 54, 82]. Prior research has explored the correlation between daily activities and mental health by employing statistical and machine learning techniques to scrutinize time-series data during [45, 68] and before the pandemic [12, 35, 73]. However, very few of these investigations can be termed as long-term tracking studies. Our research leverages a mobile sensing app that facilitates persistent, granular data collection over years, permitting remote and cost-effective capture of continuous sensor data from devices and self-reported EMA surveys. This paper presents the findings of a substantial longitudinal multi-year study spanning nearly five years involving two cohorts of 104 and 113 undergraduate students from Dartmouth College, a four-year Ivy-league University in the United States. Our study gathered up to 20 different sensory features at varying intervals and responses to numerous surveys weekly, offering a comprehensive understanding of college life and the COVID-19 pandemic's impact on students' daily routines and mental health.

Numerous existing studies provide insights into behavioral changes during the pandemic [43, 54, 78], but few of them concentrate on the return to baseline after COVID subsides. The majority of the studies that explore the behavior and mental health after COVID subsides usually just focus on the pre-post lockdown phases or a specific outbreak period [17, 38, 81]. However, our data spans pre-pandemic, pandemic, and the return to in-person instruction, hence, we believe we are ideally positioned to undertake return to baseline analysis. Our research investigates gender-based differences in students and the behavioral discrepancies between their first and last terms at university. We also explore the implications of changing living arrangements during COVID. While several studies examine changes in living arrangements concerning their influence in the spread of the COVID-19 virus [8, 34], they do not focus on the differences between on-campus and off-campus students. It should also be noted that there are more studies that investigate the impact of changing living arrangement and remote learning on students' behavior and mental health, although they primarily use self-reports only [5, 49]. Our research offers invaluable insights into the relationship between daily activities and mental health during the COVID-19 pandemic, informing the development of effective mental health support interventions for students. While similar to our longitudinal data, Xu et al. [84, 85] conducted a four-year study of 497 students using mobile sensing and surveys; each student participated in the study for a 10-week period each year. In contrast, our study maintains a consistent participant group throughout the four years. While we experienced some participant dropout—an expected occurrence in long-term longitudinal studies—we examined the behavioral and mental health changes of the same students from their college entrance to graduation. This continuity offers a more comprehensive understanding of how students' behavior and mental health evolve during their college experience, particularly in the context of a global pandemic.

3 COLLEGE EXPERIENCE STUDY

In this section, we discuss the details of our study methodology, including its design, collected passive signals and Ecological Momentary Assessments (EMAs).

3.1 Study Design

In our study, we employed a mixed-methods approach, specifically utilizing a convergent design with a supplementary qualitative component. This design enabled us to analyze the behavioral sensing data and self-reports quantitatively, while concurrently conducting interviews with fifteen undergraduate students to provide context and clarify the quantitative findings. By combining the behavioral data and EMAs with qualitative insights from students' experiences, we gained a more comprehensive understanding of the results. Our methodology comprised two phases. In the first phase, we collected mobile sensing data, EMA responses, and a variety of other validated psychometric measures from students over a four-year period. In the second phase, we sought to understand the perceptions and experiences of fifteen students by conducting in-depth interviews. It is important to note that these students, who graduated in 2022 and 2023, were not part of the main study sample. They were selected for their ability to provide valuable contextual information. This streamlined process allowed for efficient data collection and facilitated an interpretive analysis of diverse findings. Incorporating interpretation was crucial in our study, as it offered insightful context and explanations to better understand the relationships between the quantitative results and the students' experiences.

Table 1. Participant Demographics: The table presented provides a detailed breakdown of the demographic composition of the students involved in our study

Category	**Count**	**Percentage**
Sex		
Female	129	69.35%
Male	57	30.65%
Race		
White	110	59.14%
Asian	43	23.12%
Black or African American	6	3.23%
American Indian/Alaska Native	6	3.23%
More than one race	15	8.05%
Not reported	6	3.23%

3.2 Recruitment

In this study, a sample of 217 first-year undergraduate students from Dartmouth College in the United States were recruited. The university registrar sent all incoming admitted students information about the four-year study. The recruitment process took place from August 2017 (cohort 2021) to November 2018 (cohort 2022), with students enrolling and providing consent at the beginning of their first academic year. However, two participants were excluded from the dataset and subsequent analyses due to one participant's mobile phone incompatibility with the sensing application and another withdrawing within the first week of the study. Consequently, the final sample consisted of 215 participants, of which 146 (67.8%) were female. The age range for all participants at the time of enrollment was 18 to 22 years. The study was approved by Dartmouth College's Institutional Review Board (IRB). Potential participants were initially asked to complete an online survey to determine their eligibility for the study. Upon meeting the eligibility requirements (such as, possession of an Android or iOS smartphone compatible with the study's mobile application) and expressing an interest in participation, individuals proceeded to complete a series of online surveys. Subsequently, students were asked to install our data collection application on their Android or Apple smartphone. The mobile sensing application collected two types of data: sensing data to capture users' behavior and EMA data to measure mental health. Participants received compensation for their weekly EMA responses at a rate of 10 dollars per week. Note that they did not receive any additional compensation for the passive sensing data collection.

3.3 Demographics

Out of the 215 participants, 186 students have at least three years of data. Table 1 shows the demographics of the 186 students used in our analysis. The majority (69.35%, N=129) of our participants identify as females. In terms of race, 59.14% (N=110) are White, 23.12% (N=43) are Asians, 3.23% (N=6) are Black or African American, 3.23% (N=6) are American Indian/Alaska Native and 8.05% (N=15) belong to more than one race.

Table 2. Features: The table below lists the passive sensing data we collect from the participants.

Category	**Features**
Physical Activity	Walking / sedentary (still) / biking / running duration duration in vehicle, number of steps
Mobility & Semantic Locations	Distance travelled, time spent at home, workout places, study places, social places, dorms, others' dorm, greek houses, max distance from campus, number of locations visited
Phone usage	Number of phone locks & unlocks, duration of phone unlock
Audio Plays	Number of audio plays, duration of audio plays
Sleep	Sleep duration, sleep start time, sleep end time

3.4 Sensor Data

Our mobile application taps into various sensors embedded within smartphones to record data, and we further compute derivative features from this raw data. Below, we offer an overview of the types of features we gather. For a complete list, please refer to Table 2.

Physical Activity. Our application utilizes the Activity Recognition API available on both iOS and Android devices to record the physical activities of the participants.

Mobility and Semantic Locations. By employing GPS data, we gather information on students' mobility, which allows us to compute features such as the total distance travelled and the number of unique places visited. Furthermore, we assign semantic labels to various locations on campus (e.g., gym, study place, dorm, social places, cafeteria, classrooms, etc.). Since most buildings on the university campus serve a single purpose, this enables us to analyze student behavior in different contexts.

Phone Usage. As an indicator of screen-time and overall phone usage, our application tracks the number of times students lock and unlock their phones. We compute the total number of phone lock and unlock events and the average duration between these events.

Audio Plays. Our sensing application monitors whether any audio session, encompassing any audio-based media like music or video, is active on the phone. We compute the number of audio plays and the total duration of audio play.

Sleep. We estimate bedtime, wake-up time, and sleep duration using the method described in [14, 73]. Please note that the sleep data has a measurement error of approximately +/- 32 minutes.

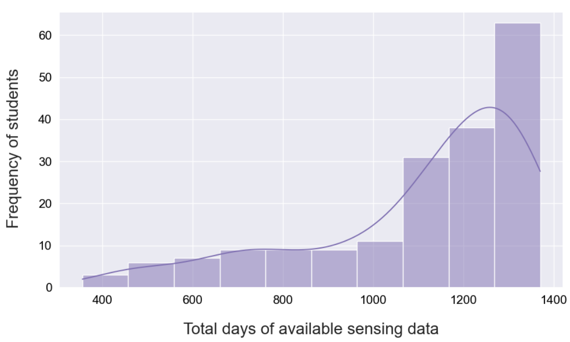
For analytical and interpretive ease, we categorize each of these feature groups into four periods: epoch 0 (overall, encompassing a full 24-hour period), epoch 1 (morning, 12 am to 9 am), epoch 2 (day, 9 am to 6 pm), and epoch 3 (night, 6 pm to 12 am).

3.5 Ecological Momentary Assessments

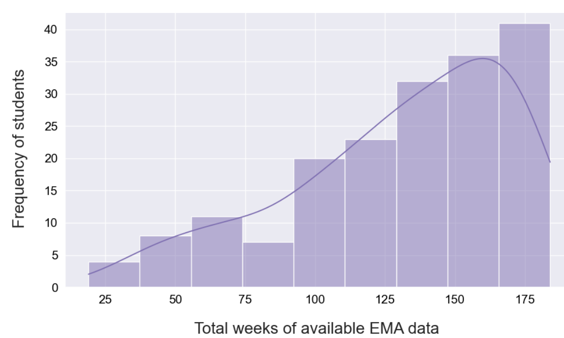
We collect weekly EMA self-reported surveys from the participating students. The categories of responses are shown in Table 3. We use the EMA responses to define the mental health outcomes in subsequent analysis. Note that the first two questions of Patient Health Questionnaire-4 (PHQ4) are used as anxiety questions, and the last two capture depression.

3.6 Data Completeness, Compliance and Challenges

We assess the consistency and quality of our data in this section, focusing on two main components: sensor data collected from smartphones and responses to EMAs. For the sensor data, we calculate a completeness ratio, representing the percentage of days with recorded data relative to each participant's total study duration. The EMA data is evaluated on a weekly basis, comparing the number of weeks a student responded to the EMA against their overall study participation duration. Additionally, we examine the missing rates for each feature to confirm there are no substantial data gaps.



(a) Sensing data availability



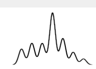
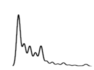
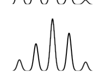
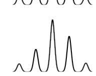
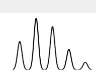
(b) EMA responses availability

Fig. 1. Data Availability: Figures (a) and (b) display sensing data and EMA self-reports availability, respectively. The x-axis represents days of sensing data in (a) and weeks of EMA responses in (b); the y-axis indicates the number of students in both. The smooth line in each plot shows the kernel density estimate, with a trend towards data availability for a higher number of days and weeks, approximating the 4-year study duration.

Figure 1a illustrates the distribution of the average number of days for which we have available sensing data per participant. On average, we have sensing data available for 1104 days per participant. This translates to an average completeness rate of 90%, implying that for each participant, we have data for 90% of the days from their enrollment date to their exit from the study. Similarly, Figure 1b showcases the distribution of the average number of weeks for which we have EMA responses. On average, participants have responded to the EMA for 130 weeks, yielding an average EMA completeness score of 75%. Both plots reveal that the data availability is skewed towards a higher number of days and weeks. The smoothed line on the plot, representing the kernel density estimate, peaks at one point, indicating that we have unimodal data with one common value for days and weeks. In terms of missing data, we find an average per-user missing value rate of approximately 20% for location-based features, 12% for audio play-based features, and 22% for steps. We have no missing data for physical activity features, and as the phone usage and sleep features are derived, they invariably have non-empty values.

We conducted Little's Missing Completely At Random (MCAR) test to examine the missing data mechanism in our dataset and obtained statistically insignificant result (p-value > 0.10), suggesting that most variables are likely missing completely at random (MCAR). Little's MCAR test is a statistical method to determine whether missing data in a dataset is completely random. We also explored the possibility of data being missing at random (MAR) for specific variables by calculating correlations between the binary indicator of missing values for each feature and the observed values of other features in the sensing data. MAR refers to a type of missing data mechanism

Table 3. Self-Report Measures: The following table enumerates the EMA self-reports collected from participants, delivered randomly once per week through our mobile application.

Self-reports	**Median**	**Std.**	**Range**	**Distribution**
COVID				
How concerned are you about COVID-19?	4.00	0.49	1-7	
Patient Health Questionnaire-4 (PHQ4)				
Over the last 2 weeks, how often have you been bothered by the following problems? Feeling nervous, anxious or on edge; Not being able to stop or control worrying; Feeling down, depressed or hopeless; Little interest or pleasure in doing things;	2.00	2.06	0-12	
Self Esteem				
Right now, I worry about what other people think of me.	2.00	0.81	1-5	
Right now, I am pleased with my appearance.	3.00	0.77	1-5	
Right now, I feel as smart as others.	3.00	0.77	1-5	
Social Level				
Have you spent most of your time alone or with others today?	3.00	0.69	1-5	
Stress				
Are you feeling stressed now?	2.00	0.66	1-5	

where the missingness is related to observed variables but not to missing values themselves. Our findings reveal that several location-based features have strong correlations (ρ-value $\geq \pm 0.40$) with missingness indicators, suggesting that location-based features could be MAR rather than MCAR. We further examined missing values using a nullity correlation matrix. A nullity correlation matrix summarizes the relationships between missing data patterns for all pairs of variables in a dataset, where nullity correlation quantifies the strength and direction of the association between missingness patterns for two variables. Figure 2 visualizes the relationships between missing values across the behavioral sensing data, showing a strong association (ρ-value ≥ 0.7) across most location-based features. This result aligns with the notion that location-based features might be MAR, implying a systematic pattern of missingness across these features. Various factors could contribute to this finding, such as participants turning off their GPS or location services on their phones due to privacy concerns or battery usage. Alternatively, missingness in location-based features might relate to the quality of GPS signals or the frequency of data collection, influenced by participants' behaviors or characteristics. While we lack complete demographic information and other characteristics of the participants, we further analyzed whether students' demographics play a role in the missingness of each feature. We calculated the missing rate per feature for each participant and used the Kruskal-Wallis test (a non-parametric test to compare independent groups) to compare the missing rates for different demographic groups (race, gender) and cohorts (2021 and 2022) for each feature. Our analysis revealed no significant difference in missingness for features across race and cohorts. However, we found a statistically significant difference (p-value < 0.10) in missingness for the audio plays feature between genders – males had a higher missing rate (18%) compared to females (8%).

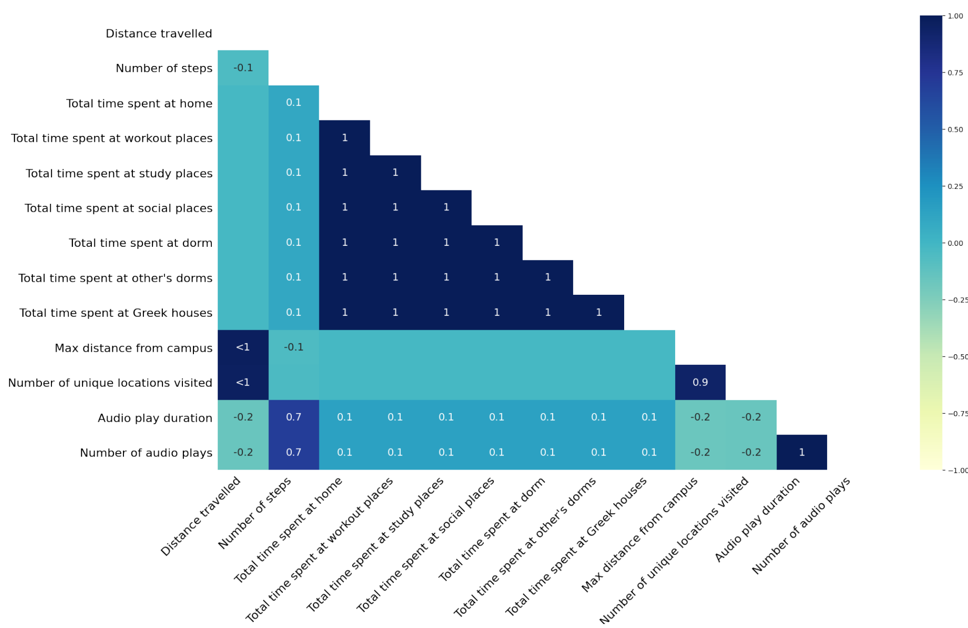


Fig. 2. Nullity Correlation: The figure illustrates the strength of correlation between the presence or absence of one feature and that of another. Most location features have a value of 1, indicating that if one location feature appears, the other almost certainly does as well. Given that location features depend on the availability of GPS data, it's logical that they would co-occur, either both being present or both absent.

In addition to the missing data patterns observed in our dataset, it is important to consider the challenges and context of conducting a four-year longitudinal study using Android and Apple phones, as these factors have influenced data collection and participant retention. Conducting such a study presents challenges due to the continuous evolution of students' phones and operating system updates. For instance, between 2017 and 2022, there were six and seven new versions of Apple iOS and Android OS, respectively. Pushing these updates often relied on cellular providers and required modifying and updating our app to ensure proper functionality. In some cases, these updates caused conflicts between the hardware or software, leading to students dropping out of the study. Moreover, even when updates went smoothly, data loss occurred as students did not immediately reinstall the app update after receiving the new OS. Consequently, compliance rates for completing the four-year study were affected by retention issues and technological updates. This situation created an ongoing race for app developers to stay ahead of significant OS changes before they were pushed by providers. One notable change was introduced in Android 10 (released on September 3, 2019), which imposed restrictions on when apps could initiate activities while running in the background. This change had a significant impact on our passive sensing application. Typically, passive sensing apps push APIs, phones, and platforms to their limits, balancing low-power sensing, machine learning inference, and cloud integration while considering energy consumption. Taking these factors into account, out of the initial study size of N=215 participants, a total of N=123 students completed the four-year study. Specifically, we have the sensing data available for N=123 participants up until their graduation, consisting of 62 students from the 2021 cohort and 61 students from the 2022 cohort.

4 FOUR YEARS OF COLLEGE LIFE

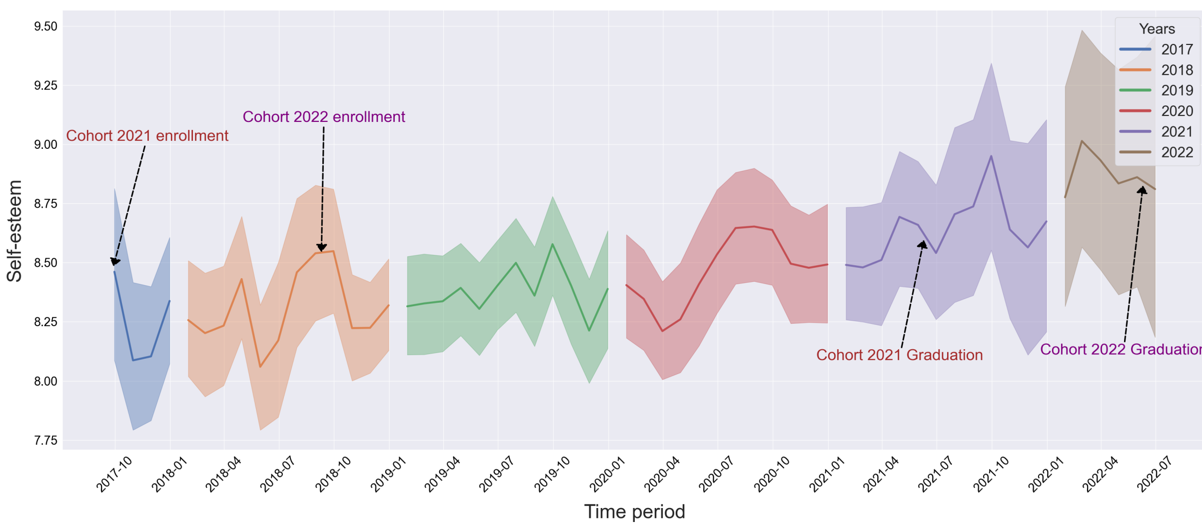


Fig. 3. Temporal Progression of Self-Esteem: The figure shows participants' self-esteem over the four-year study. The x-axis denotes the date, the y-axis the self-esteem score, and different colors represent different years. The smooth line indicates the average self-esteem score, with shaded regions displaying the 95% confidence intervals (calculated using bootstrapping). The data implies a gradual increase in participants' self-esteem throughout the study.

Dartmouth College is located in a small college town in the United States and operates on a quarter system consisting of four ten-week terms: fall (mid-September to late November), winter (January to early March), spring (late March to early June), and summer (mid-June to early September). Throughout the academic year, undergraduates enjoy winter break, spring break, and summer break. Our study, initiated in 2017, enrolled two cohorts of students in a rolling fashion. Cohort 2021 (students graduating in Spring 2021) was enrolled in September 2017, followed by Cohort 2022 (students graduating in Spring 2022) in September 2018. We collected data from each cohort over four years, from their initial days on campus until graduation. We supplemented this data with regular self-reports on students' mental health and wellbeing through self-reported EMAs. Figure 3 illustrates one such self-reported data point (i.e., self-esteem) highlighting enrollment and study exit dates for both cohorts. It is evident that students experience a sudden drop in self-esteem upon joining the university, likely due to the challenging transition from high school to college-level academics, peer pressure, or homesickness. However, over the years, their self-esteem gradually increases, reaching higher levels by the time they graduate. In addition to EMAs and mobile sensing data, we administered several surveys at various time points during the study (every six months to a year). This extensive longitudinal study provided valuable insights into students' behavior and mental health throughout their entire college experience.

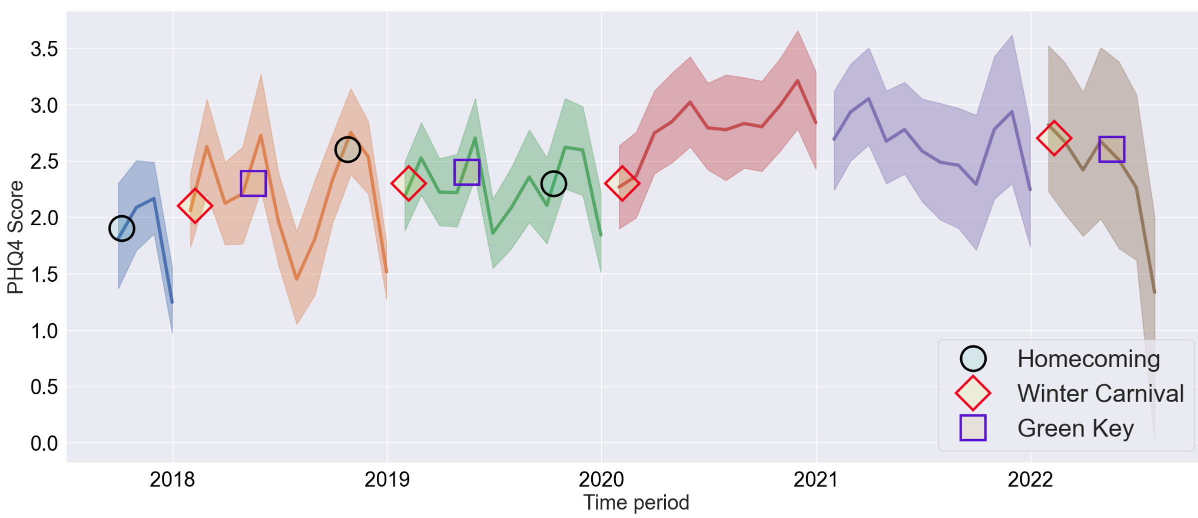


Fig. 4. Campus Events and PHQ4 Scores: The figure shows participants' PHQ4 scores throughout the year, highlighting three significant campus events. The x-axis represents time, and the y-axis the PHQ4 score. Circular, diamond, and square markers denote different festivals: Homecoming (fall), Winter Carnival (winter), and Green Key (spring), respectively. A trend of lower PHQ4 scores around these events may suggest their calming effect on students' mental health, with lower scores indicating less anxiety and depression.

Each term is characterized by a unique array of events and festivals that reflect the changing seasons and academic milestones. Fall term is characterized by *Homecoming*, a time when students, alumni, and faculty unite to honor the college's rich history and traditions through various engaging events. Winter term brings the eagerly anticipated *Winter Carnival*, a weekend-long tradition that assembles the community to celebrate and enjoy the snow-filled season. As spring term arrives, the campus blossoms with vibrant flora and warmer weather, setting the stage for *Green Key*, an annual spring event featuring a diverse array of social, cultural, and athletic activities. These events are illustrated in our data presented in Figure 4. The x-axis represents the time period, while the y-axis displays the PHQ4 score. As detailed in Table 3 the PHQ4 (Patient Health Questionnaire-4) is a brief four-item questionnaire that students self-report once per week that is used to screen for depression and anxiety. The PHQ4 is based on the PHQ9 and the Generalized Anxiety Disorder-7 (GAD7). The PHQ4 has two subscales: the PHQ2 for depression and the GAD2 for anxiety. Each item is scored on a 4-point Likert scale that ranges from 0 (not at all) to 3 (nearly every day). The total PHQ4 score ranges from 0 to 12, and total PHQ2 and GAD2 scores range from 0 to 6. Higher scores denote greater levels of depression and anxiety. Results shown in Figure 4 suggest that these campus-wide festive events typically have a soothing effect on students, as their self-reported PHQ scores tend to be lower around the time of these events. There are some periods where the PHQ4 score, although not at its peak, is trending high. Especially in the year 2019 starting with Homecoming and a few events in 2018 as well.

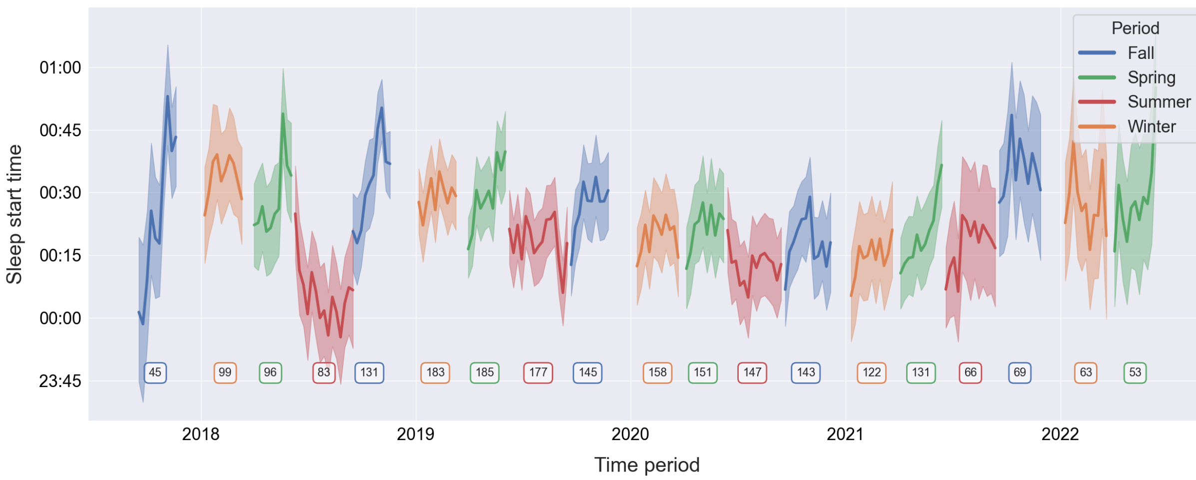
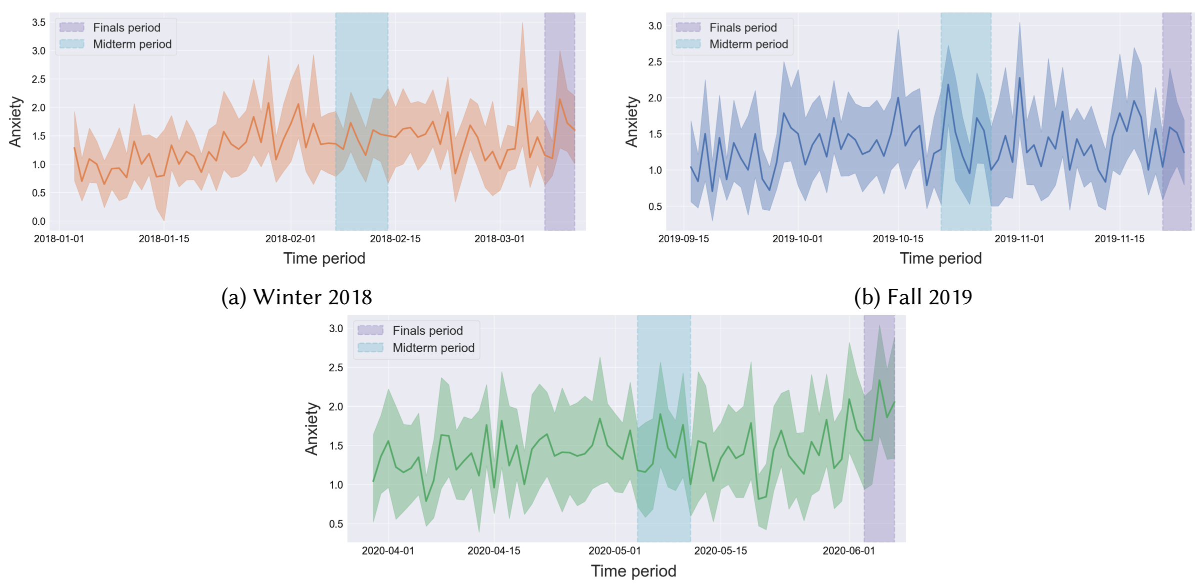


Fig. 5. Sleep Initiation Patterns: The figure shows students' typical sleep initiation times, with different terms in various colors. Rectangles at the bottom indicate participant counts with sleep data per term. Summers feature earlier bedtimes, diminishing as the year progresses. The x-axis represents the time period, and the y-axis denotes sleep start time.

The interviewed students suggested that the workload during these events could be the cause of the stress, as students are torn between enjoying the festivities or completing their work. The interviewed students pointed out that they often used to spend the weekends before these events doing their work to enjoy the festivities fully. Some students also shared that the increasing trend in PHQ4 in certain cases might be due to 'frat ban'. First year students are not allowed to take part in fraternities and sororities of the college. Homecoming of the next fall (i.e., sophomore year) is when they are allowed to go into Greek spaces. Note, all partying/drinking events happens only at the Greek spaces on campus. So a few students shared that until that point, it is just first year students hanging out with each other and having closed groups. *"But then after Homecoming, students have more access to the whole campus and that is when you start feeling like excluded from things because certain people go to certain places and not everyone can join. I think the consensus around the school is that Greek space is exclusive sometimes. So people start getting spread out and they might feel left out when others are doing something, increasing their stress."*, said one of the students we interviewed. Another student added a different perspective, *"I am in a sorority and once I joined, I felt like it is nice to have things planned out for you, especially if you are a first year and you do not know where to go because it is your first festival, but then that kind of adds stress as I feel like I should be having fun because everyone tells me that it is fun"*. Some students offered a different perspective on the impact of 'frat ban' lifting. They mentioned an increase in overall mood and a sense of becoming a full-fledged member of the campus community, which contributed positively to their college experience. They reflected that the lifting of the ban often coincided with the onset of exam periods, potentially complicating the interpretation of the observed spike in stress levels. Interviewed students also noted that while these events may change mental health at the moment, they may not have as much of an impact in the long run once students have experienced them.

In this extensive longitudinal study, we not only capture self-reported metrics from the students but also collect their behavioral data using our mobile app. Figure 5 illustrates one behavior we examined: the sleep start times of students across four academic years, specifically during the fall, winter, spring, and summer terms. The plot shows that sleep start times frequently shift within and across terms. Changes within a term may be influenced by significant milestones, such as mid-term and final exams. Across terms, we consistently observe that students tend to go to bed earlier during the summer, particularly in the first summer. This trend may be attributed to it being their initial summer break when many students are not yet engaged in internships or other activities. During an interview, one student mentioned noticing a more balanced sleep schedule, especially during the summer break, attributing it to being at home and having fewer obligations to attend parties or socialize as they did on campus. However, the difference in sleep start times diminishes in the summer of 2019. At Dartmouth, nearly all second-year (sophomore) students are required to enroll and remain on campus during the summer term, participating in various academic, social, and outdoor activities (the second-year summer term is referred to as sophomore summer throughout the manuscript). This requirement could contribute to the narrowing gap in sleep start times during this period. Additionally, the plot shows peaks in sleep start times towards the end and sometimes the middle of terms, potentially indicating mid-term and final exam periods. These peaks underscore the fluctuations in students' sleep patterns as they navigate the academic calendar and its associated demands. Note that the numbers at the bottom of the plot indicate the number of students for whom data was available



(a) Winter 2018

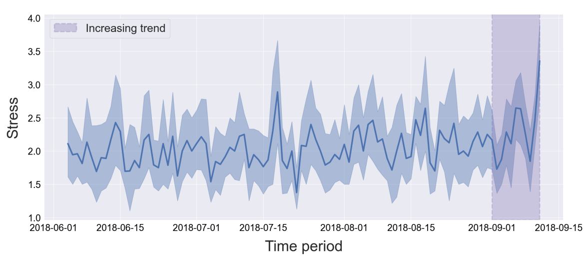
(b) Fall 2019

(c) Spring 2020

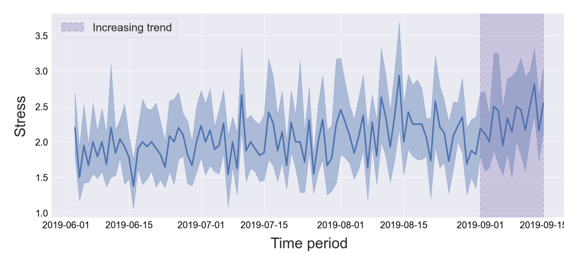
Fig. 6. Student Anxiety Patterns: The figure displays students' anxiety levels across three terms: (a) Winter 2018, (b) Fall 2019, and (c) Spring 2020, highlighting mid-term weeks and final week as critical academic periods with heightened anxiety. The x-axis indicates the time period, and the y-axis represents the anxiety score.

during the respective terms. One important factor to consider in our study is Dartmouth's quarter system, which offers students the flexibility to take terms off. This flexibility impacted our study, as some students went on study abroad programs or returned home, often choosing to disable the app during these periods. Participation in the study was entirely voluntary, allowing them the freedom to do so. Additionally, we observed a decrease in participation following 2020, which could potentially be attributed to circumstances related to the COVID-19 pandemic.

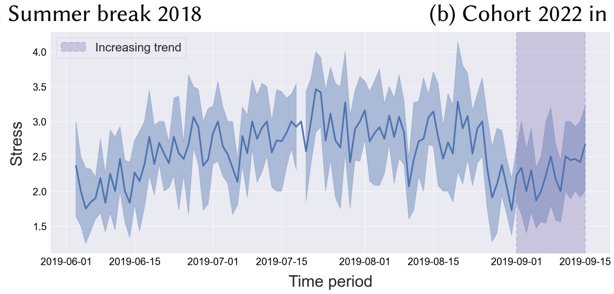
In Figure 6, we show the anxiety level of students in three different terms: Winter 2018, Fall 2019 and Spring 2020, with the exam periods highlighted. We observe that throughout the academic terms, students' self-reported anxiety levels fluctuated, with noticeable increases typically occurring during midterm and final exam periods. It is evident that anxiety levels subside once the examination period is over, allowing students to experience temporary relief from academic stress. As Dartmouth does not follow a conventional semester system, students do not have a fixed midterm examination schedule. Instead, they face the unpredictable academic schedule of midterms appearing at any time, ranging from as early as Week 3 to as late as Week 7. Consequently, the other peaks in the plot could also be attributed to this scattered exam schedule. However, midterms most often transpire during Weeks 6 or 7, which is why we have chosen to highlight these weeks in the plot. One of the students we interviewed stated that they were more stressed during midterms than finals, especially as a first year. They mentioned, *"That was one of my first college exams, and I was not really sure what to think and I would be worried that if I bomb this test, there goes my grade. By the finals period, I knew what my grade was like and I would be less worried."* The final exam period typically takes place during the last two weeks of the term, leading to a heightened level of anxiety among students as they prepare for and complete their final assessments. In addition,



(a) Cohort 2021 in Summer break 2018



(b) Cohort 2022 in Summer break 2019



(c) Cohort 2021 in Sophomore summer 2019

Fig. 7. Student Stress in Summer: The figure shows stress levels of student cohorts during summers (a) Cohort 2021 in 2018 (summer break), (b) Cohort 2022 in 2019 (summer break), and (c) Cohort 2021 in 2019 (sophomore summer term i.e., second year summer). Separate cohort analysis ensures accurate representation, as second-year students stay on campus. The shaded area reveals increased stress towards summer's end and fall term onset. The x-axis represents the time period, and the y-axis indicates the stress score.

some final exams may be in the form of projects, which require students to start their preparations early, leading to increased anxiety in the lead-up to the final submission deadline. This early-onset stress can contribute to heightened anxiety levels throughout the term, even before the official final exam period begins. This pattern emphasizes the significant impact that exam-related stress, including both traditional exams and project-based assessments, can have on students' mental well-being and underscores the importance of understanding and addressing the causes of anxiety throughout the academic term.

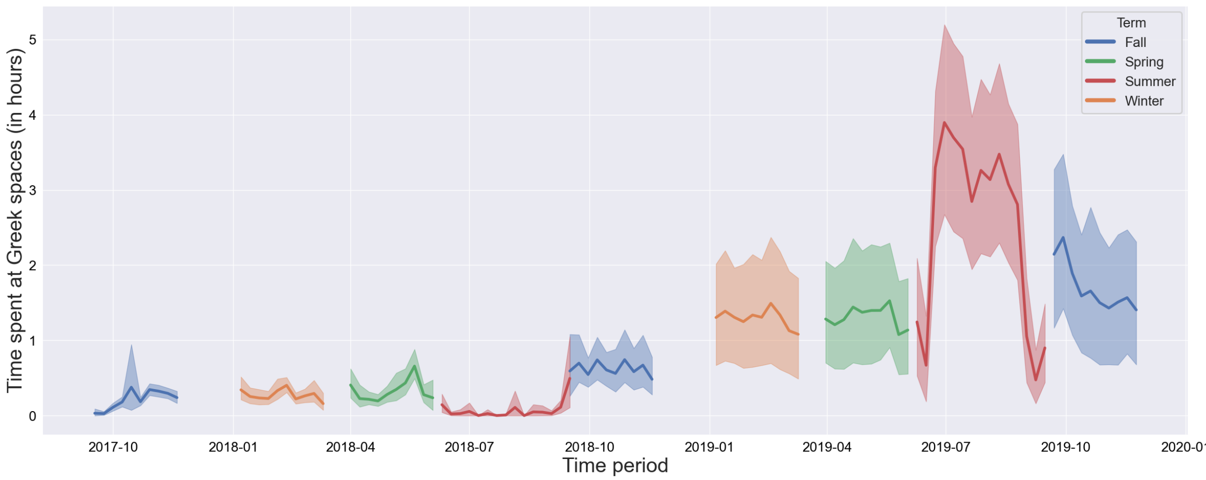


Fig. 8. Partying/Social Patterns: The figure displays time spent by the 2021 Cohort at Greek spaces (fraternities and sororities) during their first two years. The x-axis indicates the date, and the y-axis represents the duration in Greek spaces, in hours. Different colors signify academic terms. First-year students are restricted from attending Greek spaces, lifting in their second year's fall term. A peak in attendance occurs during the sophomore summer term, reflecting heightened social activity or prime party season.

Figure 7 illustrates student stress levels over two summers, segmented by cohort. The first two plots depict the first summer break for the 2021 (Summer 2018) and 2022 (Summer 2019) cohorts. In contrast, the lower figure, Figure 7c, highlights the increased stress during the sophomore summer term for the 2021 cohort, which can be attributed to mandatory on-campus activities and classes (recall that second year students have to stay on campus and enroll in classes during the summer). The highlighted regions in each graph indicate a surge in stress towards the end of summer, typically a few weeks prior to the Fall term commencement. This is an interesting phenomenon. This trend is particularly pronounced for the 2021 cohort in Summer 2018 (Figure 7a), where stress levels skyrocket near the summer break's conclusion. The anticipation of returning to campus, facing academic pressures, and transitioning from a relaxed home environment to a more structured campus life, may contribute to this late summer stress spike. This observation was reinforced by our student interviews, where they noted a surge in commitments even before the official start of the fall classes. One student reflected, *"Some sports and clubs have to begin preparing for the fall term a few weeks in advance. We also have to get ready packing for the rest of the year and preparing travel arrangements, which is stressful sometimes."* This suggests that both academic and logistical pressures are potential stressors as students approach the fall term. Several students highlighted that the summer break after the first year of college stands out from subsequent breaks. They pointed out that during the first summer break time, they often reconnect with high school friends who also return home. As the break ends, there's a sense of leaving behind the comfort and familiarity of these long-standing friendships to return to the relatively newer college environment. This transition, they suggested, could contribute to the spike in stress levels experienced at this time.

Despite the escalating stress levels observed in the 2021 cohort throughout their sophomore summer, as shown in Figure 7c, this period paradoxically emerges as a peak party season for students. Figure 8 demonstrates a substantial surge in the time students spend in Greek spaces on campus during this second summer term. Sophomore summer is the biggest partying scene across all terms of college. This is the summer when only the second year students are on campus. The teaching load is considered to be lighter than other terms. According to the students we interviewed, it's typical during the sophomore summer term to enroll in only two courses, resulting in ample free time. Additionally, a campus-wide pong tournament (which often takes place in Greek spaces) is a popular summer event, with many students dedicating a significant amount of time to practice and participate. Students are clearly much more social as shown in the data. These Greek spaces, incorporating fraternities and sororities, serve as the primary venues for social events and drinking parties on campus. Note, all drinking parties are based in the Greek spaces on campus and are open to all students across campus. Thus, an extended time spent in these areas likely reflects enhanced engagement in social activities, essentially, more partying. However, a noteworthy observation from the figure is the negligible presence of students in Greek spaces until Fall 2018. This can be explained by a college-wide policy that bars first-year students, in our case, the 2021 cohort, from entering these Greek spaces (i.e., 'frat ban'). Once this restriction is lifted after the first year, there is a considerable uptick in Greek space attendance during subsequent terms.

4.1 COVID Strikes

The COVID-19 pandemic struck midway through our study in March 2020, affecting Cohorts 2021 and 2022, set to graduate one and two years post-outbreak, respectively. Dartmouth College swiftly implemented CDC-recommended social distancing policies. Our application persisted in data collection throughout, enabling insights into the pandemic's impact on student behavior and mental health. Furthermore, in order to enrich our analysis, we conducted interviews with students from the 2020-2021 and 2021-2022 academic years. These discussions, combined with the quantitative insights from the data, deepen our comprehension of the challenges students faced during this unprecedented time and underscore the significance of adaptability and resilience in the educational landscape.

Figure 9 shows students' COVID-19 concerns from March 2020 to May 2022, with key time points corresponding to campus policy changes. Concerns peaked during the initial outbreak in early 2020, when remote teaching began. In May 2020, daily symptom screening started, and mask-wearing was enforced on campus. In September 2020, Dartmouth brought back more than half of the undergraduates, with safety measures like testing and face-covering policies, resulting in a rise in concern. In October 2020, a COVID-19 uptick occurred in the college town, but the on-campus community remained unaffected. Biweekly testing started in January 2021, and vaccinations became available in April 2021, causing a decline in student concern. However, the Delta variant in July-August and Omicron variant in December-January increased concerns. With vaccination rates rising and restrictions easing, concerns gradually decreased in 2022, and the university removed the indoor mask requirement by March.

Interviews with students who graduated in 2022 shed light on their experiences during the COVID-19 pandemic at the university. They reminisced about the summer of 2021, when a short-lived removal of the mask mandate was followed by a reinstatement due to a significant outbreak. By Fall 2021, in-person classes resumed without a mask mandate, offering students a sense of relief as the university seemed to be handling the pandemic effectively. Although some professors still held occasional Zoom sessions in Winter 2022, most classes took place in person. Dartmouth transitioned to remote instruction and pass/fail grading during the pandemic, providing students with

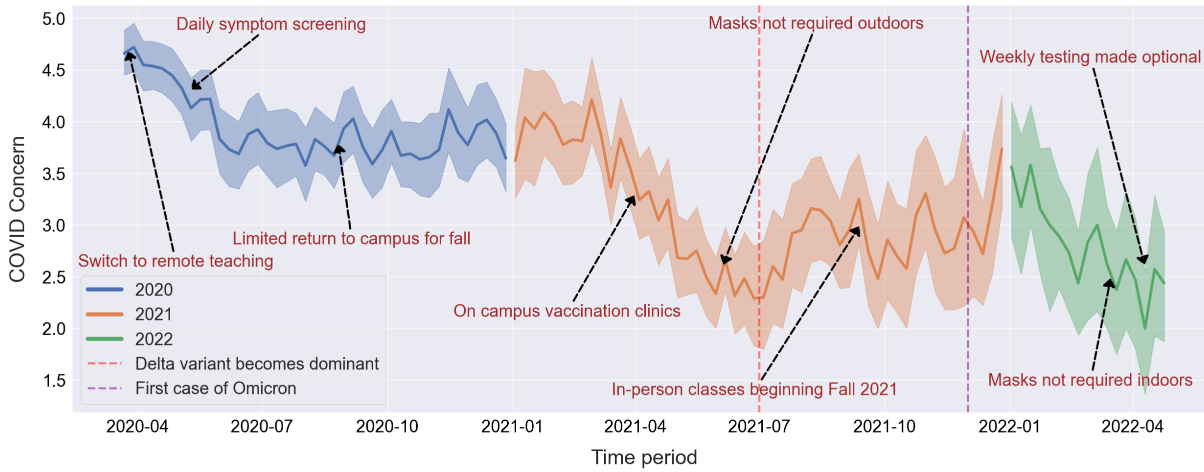


Fig. 9. COVID-19 Concern Over Time: This figure charts the fluctuations in students' concern regarding the COVID-19 pandemic over various time periods. The x-axis represents the time period, while the y-axis quantifies the COVID-19 concern scores. Different colors correspond to different years of the pandemic. The chart also accentuates several notable campus events. It appears that the level of COVID-19 concern among students responds to changes on campus, outbreaks, and the emergence of new variants (which is indicated in the plot with the vertical dashed lines).

greater flexibility in selecting their courses. This encouraged many students to take more challenging classes to improve their GPAs, leading to unique challenges for both students and professors as they adapted to online learning. However, the pass/fail grading system also induced stress for some students. One graduate shared their experience: *"I think students were stressed up in the beginning and then we had pass/fail grading, so we laid up a little to the point where my friends and I sometimes, took a little easier on assignments. Professors mentioned that if we had a certain grade, we would not need to take the finals. And by the end of the term, during the finals period, we were just figuring out that we do not have the required grade. So we had to take the finals and also had to do well in the final in order to actually pass"*. They added, *"It was an ambiguous term and we did not know what accounted for pass or fail. We realized it would look a lot worse to have fail from a class rather than a B or a C"*. While the majority of students appreciated the pass/fail grading system, some expressed dissatisfaction, believing that the lack of differentiation in grading reduced their incentive to work hard.

In summary, Dartmouth faced unprecedented challenges during the COVID-19 pandemic. By implementing policy changes, safety measures, and academic adjustments, it managed the crisis and started returning to normalcy. As the institution moves beyond the pandemic era, it is important to observe the ongoing impact on student life and academic performance and apply the lessons learned to create a stronger, more adaptive educational environment for future generations. This paper focuses on both pre-pandemic aspects and the significant influence of COVID-19 on students' lives during the study period.

5 FIRST VS FINAL TERM: BEHAVIORAL SHIFTS IN UNDERGRADUATE STUDENTS

The first term of college is a critical period during which students adapt to a new academic environment, establish social connections, and develop essential skills to help them succeed. In contrast, the final term is when students solidify decisions about future careers and/or further studies. Final-term students likely experience substantial growth in their values, beliefs, and sense of self, which may be evident in the behavioral data and self-reports we obtain from them. In this section, we compare the first term on campus, upon entering college as undergraduates,

Table 4. First Term vs. Final Term: The table below presents the results from a Wilcoxon signed-rank test comparison between the first and final terms of two cohorts - Cohort 2020 and Cohort 2021. The significance levels, adjusted using the Benjamini/Hochberg FDR correction, are denoted as follows: (*** $p < .01$, ** $.01 \leq p < .05$, * $.05 \leq p \leq .10$). The 'first' and 'final' term columns represent the median scores for each period, while 'W' refers to the W-value, which represents the expected proportion of false positives among the statistically significant results. The epochs (0-3) correspond to different periods within a day: epoch 0 is the 24-hour average, epoch 1 is morning (12 am - 9 am), epoch 2 is daytime (9 am - 6 pm), and epoch 3 is nighttime (6 pm - 12 am). Note: All duration-related features, except 'total time spent at home,' are measured in minutes. The 'total time spent at home' feature is quantified in hours, while distance-based features are recorded in miles.

Behavior/Self-report	**Cohort 2021**			**Cohort 2022**		
	First term	Final term	W	First term	Final term	W
Duration in vehicle [Epoch 0]	20.12	29.92	510***	16.08	33.01	243***
Duration in vehicle [Epoch 2]	11.35	19.40	351***	8.43	18.93	214***
Duration in vehicle [Epoch 3]	5.75	8.25	684	4.63	10.91	333***
Still duration [Epoch 2]	415.56	431.64	474***	421.08	393.75	328***
Walking duration [Epoch 2]	84.28	90.22	786	98.52	107.72	385***
Total time spent at home	10.68	11.39	658	11.01	7.57	216***
Distance travelled [Epoch 0]	27.86	12.62	386***	26.98	27.76	351
Distance travelled [Epoch 3]	6.24	2.87	353***	3.64	6.43	279
Phone unlock duration per hour spent at home	6.97	9.75	182**	8.30	7.94	304
Number of phone unlocks per hour spent at home	2.44	3.14	155***	3.35	2.74	150**
Max distance from campus [Epoch 3]	32.33	8.14	330	31.82	11.64	181**
Number of locations visited [Epoch 0]	4.96	2.05	8***	4.88	3.53	73***
Audio play duration [Epoch 0]	114.90	161.64	236***	108.21	165.37	216***
Audio play duration [Epoch 1]	12.00	25.15	250***	18.37	31.39	363*
Audio play duration [Epoch 2]	51.50	79.38	180***	47.31	86.68	165***
Audio play duration [Epoch 3]	35.76	50.20	360**	38.49	48.04	214***
Number of audio plays [Epoch 0]	4.62	6.62	156***	4.80	5.91	208***
Phone unlock duration [Epoch 0]	139.79	212.56	247***	177.54	188.52	595
Phone unlock duration [Epoch 2]	67.07	106.32	173***	81.58	91.66	469**
Phone unlock duration [Epoch 3]	49.74	66.51	351***	66.99	62.27	755
Number of phone unlocks [Epoch 0]	69.27	91.58	300***	94.95	84.70	515*
Number of phone unlocks [Epoch 2]	33.30	47.72	240***	45.62	45.23	722
Number of phone unlocks [Epoch 3]	26.80	31.76	376***	39.34	29.00	389***
Number of steps [Epoch 0]	8844.01	6773.00	113***	9299.77	8718.46	232**
Number of steps [Epoch 1]	1091.00	658.83	202***	1007.59	792.43	284*
Number of steps [Epoch 2]	4820.68	3674.35	127***	5113.45	4772.58	347
Number of steps [Epoch 3]	2661.64	2047.74	222***	2795.32	2816.04	290*
Self esteem	8.22	8.85	409***	8.26	8.94	275***
PHQ-4	1.78	2.08	591*	2.32	2.09	537
Depression	0.50	0.92	336***	0.81	0.74	501

with the last term of their college experience, separately, based on their cohorts. Our data is non-normally distributed, which led us to choose non-parametric tests for our analysis. Specifically, we use the Wilcoxon signed-rank test, a non-parametric version of the paired t-test. This test is more appropriate for our data since we are comparing the differences in behavior between the first term and final term of the same students. We report the behaviors that are statistically significant after correcting for multiple comparisons using the False Discovery Rate (FDR) approach. FDR is a statistical measure that controls the proportion of false positives among statistically significant results in multiple comparisons testing. We list the result of comparison in Table 4.

We observed that Cohort 2021 participants (students who graduated in Spring 2021) spent more time in their vehicles during their final term, both overall and during the daytime periods. However, they were less active during the day in their final term compared to their first term. They traveled less overall and during the night, which may explain the decrease in unique locations they visited. They used their phones more frequently at home. Nevertheless, we found that they had a higher number of instances where they played media or audio, both overall and in the morning, day, and night. This trend was also observed in the number of audio plays. Additionally, both phone usage duration and the number of unlocks increased during the final term in all periods, except in the morning. We also noted that Cohort 2021 students took fewer steps during their last term compared to their first term. We observed similar findings for Cohort 2022 participants (i.e., students graduating in Spring 2022) in terms of vehicle usage—it was higher during the final term compared to their first term. They also had an increase in walking duration during the day and, unlike Cohort 2021, they were less sedentary in the day during the final term. They also spent less time at home or in their dorms. They remained closer to the campus during the night and visited fewer unique locations overall. In terms of self-reports, we found only three statistically significant variables: self-esteem, PHQ-4 score and depression score. Both cohorts had higher self-esteem in their final term compared to the first term. However, only Cohort 2021 had statistically significant difference in the first term and final term for PHQ-4 and depression score – both of which increased in their final term compared to their first term.

According to the interviews, the observed difference in behavior of the two cohorts might be attributed to COVID-19. The final term for Cohort 2021 was the spring of 2021 when the campus was still strict on gatherings and campus events, whereas for Cohort 2022, the final term was spring of 2022 when many COVID restrictions had been lifted. A lot of classes were online, and several students had pushed their major culminating experiences to spring, which the interviewees see as one of the reasons for the higher sedentariness in Cohort 2021 – they might have been just working more. They might have pushed off some of their hard classes and might have been doing their senior thesis along with it. Normally, the students explain, the last term is an easy term. One of the interviewed students said, *"in the last term, a bunch of people that we know, including us, are only taking two classes. I have so much time, I have more time than I have ever had. I do no work."* Another said, *"a lot of people I know are just spending a lot of time hanging out with each other, especially since the weather gets nice."* This supports our findings from quantitative analysis for Cohort 2022, where they are more active than their first term.

In addition, the interviewed students mentioned that there is an annual college festival (i.e. *Green Key*) that takes place in week 7 of the spring term. Due to COVID, the festival was canceled in 2021 but it took place in 2022. The students interviewed mention that this could add to why walking duration is higher for Cohort 2022. Similarly, the higher phone usage duration for Cohort 2021 could be because of remote classes as well as to connect with friends. The proximity to campus, the students mention, could be because students get nostalgic as it is their final term. They know that they are graduating and they want to see all their friends. In addition, following the week when the final exams take place, there is a dedicated senior week when the seniors hang out with their class before the graduation ceremony. So students stay closer to the campus and leave late after graduation. In terms of vehicle usage, students noted that the duration in vehicle increased for the final term compared to their first term, probably due to students not being allowed to bring cars to campus in their first year.

6 GENDER BASED DIFFERENCES IN BEHAVIORS

Our study comprises 57 students identifying as male and 129 as female. We aggregated data for each gender over the four-year period and compared them using the Kruskal-Wallis test. We chose the Kruskal-Wallis test to compare behavioral differences between the two independent groups, as our data was non-normally distributed. This non-parametric method avoids biases and inaccuracies from parametric tests that assume normality, ensuring a more accurate and robust analysis of our findings. Note that to account for multiple comparisons, we applied the False Discovery Rate (FDR) correction to the obtained p-values. The result is listed in Table 5.

Our analysis revealed several notable gender-based differences. Consistent with previous research [24, 31], male students exhibited higher levels of physical activity than their female counterparts. They walked more during all periods (overall, day, morning, and night) and recorded a higher number of steps (during the morning and night) than their female counterparts. In contrast, female students displayed higher levels of sedentary behavior. They had longer overall still durations across all periods, as well as longer still durations at social places and their own dorms. Our findings align with previous studies that have reported similar trends in gender-based sedentary behavior [27, 71]. Interestingly, our analysis also revealed differences in audio play habits and sleep patterns. Male students listened to more audio (or played more media) on their phones, both in terms of frequency and total duration (overall and during the morning). On the other hand, female students reported longer sleep durations, as well as earlier bedtimes and wake-up times compared to males. Finally, we found that female students unlocked their phones more often during the morning (epoch 1).

Table 5. Female vs. Male Participants: The table below illustrates the results from a Kruskal-Wallis test comparison between female and male participants in our study. The significance levels are adjusted using the Benjamini/Hochberg FDR correction and are denoted as: (*** $p < .01$, ** $.01 \leq p < .05$, * $.05 \leq p \leq .10$). The "Female" and "Male" columns display the median scores for each gender, and 'H' refers to the H statistic, which is a measure of the difference in median output scores between the two genders. A larger H statistic indicates a greater difference in median output scores. The epochs (0-3) delineate different periods within a day: epoch 0 signifies the 24-hour average, epoch 1 represents morning (12 am - 9 am), epoch 2 corresponds to daytime (9 am - 6 pm), and epoch 3 captures nighttime (6 pm - 12 am). Please note: All duration values, barring sleep duration, are expressed in minutes. Sleep duration is specified in hours. Both sleep start and end times are indicated in 24-hour format.

Behavior/Self-report	**Female**	**Male**	**H**
Walking duration [Epoch 0]	162.19	222.16	23.15***
Walking duration [Epoch 1]	17.51	26.05	22.79***
Walking duration [Epoch 2]	88.85	107.27	9.85***
Walking duration [Epoch 3]	56.33	79.29	27.70***
Still duration per hour spent at dorm	56.10	54.14	37.40***
Still duration per hour spent in social places	45.11	39.98	37.90***
Still duration [Epoch 0]	1222.15	1160.98	52.66***
Still duration [Epoch 1]	516.86	504.99	53.85***
Still duration[Epoch 3]	290.38	261.37	66.36***
Audio play duration [Epoch 0]	154.39	191.30	6.53**
Audio play duration [Epoch 1]	25.53	38.59	9.85***
Number of audio plays [Epoch 0]	5.75	6.85	5.55**
Number of phone unlock [Epoch 1]	9.73	10.96	6.53**
Sleep duration	7.62	7.19	5.17*
Sleep start time	23:48	00:27	15.56***
Sleep end	07:25	07:42	7.76**
Social level	3.23	3.02	10.36***
Self esteem	8.11	8.95	21.17***
COVID concern	4.16	3.66	7.76*

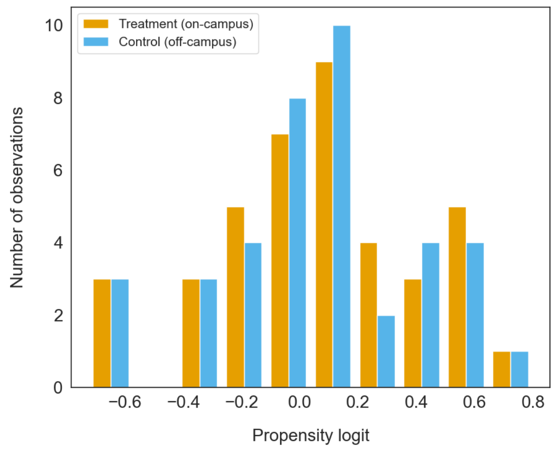
In terms of the students' self-reported surveys (i.e., EMAs), female students appeared more sociable, reporting that they spent more time with others than being alone (i.e., 'social level' question on the EMA), while male students reported higher self-esteem scores. Prior studies have shown that (based on how self-esteem is conceptualized), females tend to report lower self-esteem than males [28, 86]. Regarding the EMAs conducted during the COVID-19 pandemic, female students reported higher overall concern about COVID-19 and higher concern for themselves than their male counterparts. One of the female students we interviewed, shared her experience, *"I think that sometimes college guys are a little more negligent, if that makes sense. A bit more self-absorbed. They act like it is their world that we are just living in. Whenever there was a COVID outbreak, all of my female friends would get tested whereas the guys would always kind of go, I don't know, whatever. I actually had COVID and exposed some friends of mine. This was during my junior fall. All the girls that I told were exposed, tested as they were worried. They were continually checking up to make sure that they did not have it and take it home to their family because it was at the end of the term. But all the guys did not care and they went home without a worry in the world. So I feel like that could have been the COVID concern, even though the girls might have been hanging out with each other more, they are more aware of the impact whereas the guys might not think about it."*

While the students we interviewed could not provide any insights into the underlying reasons for other gender-based differences, our findings, particularly those related to physical activity, suggest that female students in our cohort may be at a higher risk for negative health outcomes associated with sedentary behavior. As for audio play habits and sleep patterns, the available research is inconclusive. Some studies report that female students tend to sleep longer on free days [59], while others find no difference in sleep duration [56]. There is also limited research on whether female students tend to unlock their phones more often than male students. Overall, further research is needed to fully understand gender differences in audio listening habits, sleep, and phone usage.

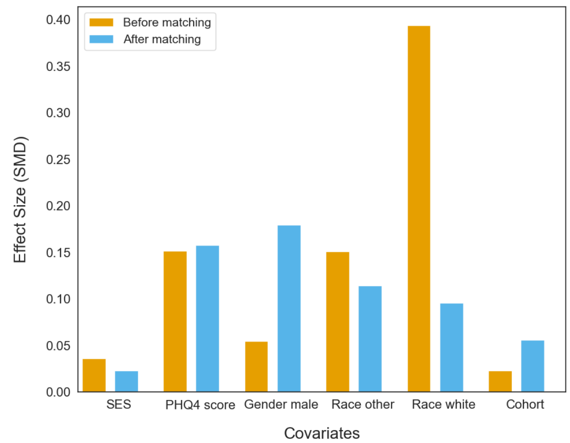
7 THE IMPACT OF CHANGING LIVING CONDITIONS

As the COVID-19 pandemic unfolded, universities had to adapt to remote learning. In this section, we aim to investigate the effects of on-campus versus off-campus living arrangements on students' behaviors and mental health during the pandemic. Dartmouth College began returning a limited number of students to campus in Fall 2020, with all students returning by Fall 2021. Consequently, we have three terms of data for comparison: Fall 2020, Winter 2021, and Spring 2021. Note that for this analysis, we use the students' GPS data to identify whether they are off-campus or on-campus. If they are within the 5 miles (ca. 8 km) radius to the campus, we term them to be on-campus, otherwise they are off-campus.

To account for potential confounding factors and non-random assignment of living arrangements, we employed propensity score matching (PSM) [61]. PSM is a statistical technique used to create comparable groups of individuals when there is no random assignment to a treatment (in this case, being on or off campus). It aims to balance the distribution of observed covariates between treatment (on-campus) and control (off-campus) groups, thus reducing bias in estimating the treatment effect. In our case, the observed covariates include race, gender, cohort, socio-economic status (collected during their enrollment), and average PHQ4 score (which is reflective of their mental health). To perform PSM, we first estimate the propensity logit scores for each student based on the set of observed covariates using logistic regression. The propensity score represents the probability of being in the treatment group, given the observed covariates. Next, we match the students in the treatment and control groups based on their propensity scores, creating a matched sample with a similar distribution of covariates in both groups. We use nearest neighbor to create a matched sample with one-to-one matching. By matching on-campus and off-campus students based on the observed covariates, we created comparable groups for analysis. We performed PSM separately for each term and assessed the balance of confounding variables using standardized mean differences (SMDs). We show the plot of SMD of the covariates before and after matching in Figure 10b and the distribution of the on-campus and off-campus group after matching in Figure 10a. We iteratively refined the matching approach until satisfactory balance was achieved, as evidenced by most SMD below 0.20 in the plot. As shown in the plot, SMDs are reduced after matching, indicating that the covariate distributions are more similar between the treated and control groups. In other words, this ensures that the matching process has successfully minimized the influence of confounding factors. After matching, we used the Kruskal-Wallis test to compare



(a) Distributions



(b) Effect sizes

Fig. 10. Propensity Score Matching: Figure (a) shows the treatment and control groups in comparison to one another post-matching. The closer the two distributions align, the more effective our matching procedure. Figure (b) presents the effect size of various covariates (calculated using Cohen's D) pre and post-matching. Cohen's D is a standardized measure of the difference between two means, expressed in standard deviation units. A decreased effect size post-matching indicates improved matching quality. While some effect sizes may increase post-matching, most of them remain below our threshold of 0.20. Both plots are specifically based on Fall 2020 data.

behaviors and mental health self-reports between on-campus and off-campus students. If a significant difference was found, we calculated the average treatment effect on the treated (ATT) for each outcome to evaluate the impact of living arrangements during the pandemic. The result is listed in Table 6.

Our findings show a general trend of significant behavioral differences between on-campus and off-campus students during Fall 2020, which diminishes over time. For fall term, our matched sample has 52 participants each in on-campus (treatment) and off-campus (control) group. In Fall 2020, on-campus students visited more locations, engaged in more audio/media play, and unlocked their phones more frequently compared to off-campus students. The ATT for still duration during morning (i.e., epoch 1) is approximately -8.97, indicating that the changes in living arrangement decreases still duration by 8.97 minutes on average for those students who returned to campus. Similarly, students on campus had an increase in walking duration during morning by 6.10 minutes on average. They also were using their phone more. In fact, on-campus students had on average, 40.08 minutes increase in phone unlock duration (overall). Being on campus also increased their phone unlock duration over the morning (epoch 1) and day (epoch 2). We observed no significant differences in self-reported mental health metrics. However, as the academic year progressed, the differences between on-campus and off-campus students became less significant. We have 42 and 21 students in each group during Winter 2021 and Spring 2021, respectively. However, by Winter 2021, fewer behavioral differences persisted, with on-campus students taking more steps overall and during different times of the day. Interestingly, we found that returning to campus during Winter 2021 decreased students' social level by 0.42 units on average. The interviewed students pointed out why that might be the case, *"There was the first outbreak in Winter 2021 and that was terrifying"*. They added, *"You could be put into isolation for a really long time even if you did not have COVID. You also could not get tested again. This is even if you were in just a general vicinity of the positive person. You were put into quarantine for two weeks. Everyone was afraid to contact trace anyone else in case they got mad at each other."* It is plausible that the outbreak could have led to the University imposing strict isolation, social distancing and testing guidelines, negatively impacting the social level of students. A few of the students mentioned that the increase in walking duration for on-campus students is probably because the only way they could have social time on-campus was by going outside. *"And so a lot of people would go for walks and just be socially distant, like six feet apart, wearing your masks or whatever the restriction was at the time"*, mentioned one of the students. Students could not go inside buildings together, but they would just walk around campus a bit more so they could be social and talk to people without getting in trouble. Students also stated that the reduction in social level during winter might have also been caused by the weather itself which meant they could not be outside as much to go for their social walks.

Table 6. On-campus vs. Off-campus: The table below demonstrates the impact of changing living arrangements during the pandemic. We initially perform propensity score matching to pair the control group (off-campus) and the treatment group (on-campus), based on observed covariates. A Kruskal-Wallis test is then used to compare these groups. The significance levels, adjusted with the Benjamini/Hochberg FDR correction, are denoted as follows: (*** $p < .01$, ** $.01 \leq p < .05$, * $.05 \leq p \leq .10$). The 'Average Effect on the Treated' (ATT) column reflects the average difference between the medians of the two groups. 'Off-campus' and 'On-campus' columns showcase the median scores for each group, while 'H' represents the H statistic. Epochs (0-3) correspond to different periods within a day: epoch 0 refers to the 24-hour average, epoch 1 to the morning (12 am - 9 am), epoch 2 to daytime (9 am - 6 pm), and epoch 3 to nighttime (6 pm - 12 am). Note: All durations, except sleep duration, are measured in minutes. Sleep duration is denoted in hours.

Term	Behavior/Self-report	Off-campus	On-campus	H	ATT
Fall 2020	Still duration [Epoch 1]	523.28	514.31	8.78**	-8.97
	Walking duration [Epoch 1]	10.35	16.46	9.82***	6.10
	Number of locations visited [Epoch 0]	1.57	2.01	15.75***	0.44
	Number of locations visited [Epoch 3]	1.22	1.45	12.30***	0.23
	Audio play duration [Epoch 0]	168.41	236.44	5.30**	68.03
	Sleep duration	7.89	7.22	11.14***	-0.66
	Number of phone unlocks [Epoch 0]	67.57	99.08	14.78***	31.51
	Phone unlock duration [Epoch 0]	191.42	231.57	5.35*	40.08
	Phone unlock duration [Epoch 1]	27.26	34.25	5.79**	7.00
	Phone unlock duration [Epoch 2]	98.87	117.56	5.36*	20.70
	Number of steps [Epoch 0]	3317.39	5831.68	21.07***	2514.29
	Number of steps [Epoch 1]	273.13	545.14	21.92***	272.02
	Number of steps [Epoch 2]	2262.82	3824.18	19.19***	1561.36
Winter 2021	Social level	3.08	2.65	5.94*	-0.42
	Number of steps [Epoch 0]	3231.81	5188.76	14.61***	1956.94
	Number of steps [Epoch 1]	278.57	456.67	6.50*	178.10
	Number of steps [Epoch 2]	2189.74	3477.32	15.00***	1287.58
Spring 2021	Number of locations visited [Epoch 3]	1.17	1.59	8.01*	0.42

Moving on, in Spring 2021, only one variable remained significantly different between groups: the number of locations visited at night, with on-campus students visiting 0.42 more locations on average. We do not have enough participants off campus to compare fall term and beyond, as the majority of the students returned to campus beginning Fall 2021. The interviewed students mention that by the end of summer and the beginning of Fall 2021, *"it was less of a big deal if you came into contact with someone with COVID. You would still go into quarantine for at least 10 days if you tested positive. But you were allowed to go home, which you were not allowed before. You were not allowed to leave campus at all."*

8 THE LASTING IMPACT OF COVID-19 ON BEHAVIORAL PATTERNS AND MENTAL HEALTH

In this section, we assess how the patterns in daily routines and mental health outcomes persist over time. In order to assess the behavioral changes over the four years, we group the students' data into three periods: Pre-COVID (i.e., data from September, 2017 to March 13, 2020; beginning of the study to the date when COVID was declared a national emergency), During-COVID (i.e., data from March 13, 2020 to September 12, 2021 when classes were remote) and Return to campus (i.e., data from September 13, 2021 to June 15, 2022 when the majority of the students were back in Dartmouth's campus for in-person instruction). We want to see if identified patterns during the COVID period persist once in-person classes begin and students return to Dartmouth or if students exhibit significant changes in their daily routines and mental health outcomes. If the patterns persist, this might suggest that the behavioral or mental health patterns identified during the COVID period have lasting effects on students. If the patterns change, it may indicate that students are adapting to the post-COVID period, and their daily routines and mental health outcomes are evolving. We used mixed effects modeling to account for the within-subject variability, repeated measures for each participant, and potential non-linear patterns in the data. Our fitted mixed effects model, shown in Equation 1, considers each behavior or mental health outcome as a dependent variable. We use the participant identifier as a random effect and we control for gender, race, and the cohort the student belongs to. We use the period (i.e., pre-COVID, during-COVID and in-person instruction) as a fixed effect. This model will provide us with the average effects of each time period on the behavior measure and mental health outcome, controlling for gender, race and cohort, while accounting for the within-subject correlation through the random intercept for each student. We can then compare the estimated coefficients for the time periods to assess if there are significant differences in behaviors between the three time periods. Note that for this analysis, we only compare the epoch 0 (i.e., overall) behavior.

$$Y_{ij} = \beta_0 + \beta_1 \cdot \mathrm{Period}_{ij} + \beta_2 \cdot \mathrm{Gender}_i + \beta_3 \cdot \mathrm{Race}_i + \beta_4 \cdot \mathrm{Cohort}_i + u_i + e_{ij} \tag{1}$$

Here, Y_{ij} is the dependent variable (behavior or mental health outcome) for student i at time j, Period_{ij} is a categorical variable representing the time period (pre-COVID, during-COVID, return to campus) for student i at time j, Gender_i, Race_i, and Cohort_i are control variables for student i, u_i is the random effect for student i (capturing the within-subject correlation), and e_{ij} is the residual error for student i at time j. In this model, β_0 is the intercept, β_1, β_2, β_3, and β_4 are the fixed effect coefficients, and the u_i term represents the random intercepts for each student. We confirmed the assumptions for our mixed-effects analysis, including a Variance Inflation Factor (VIF) < 10 for multicollinearity and assessing independence of effects. VIF is a measure of multicollinearity in statistical models, indicating how much the variance of an estimated coefficient is inflated due to collinearity. Note, most works consider VIF > 10 as an indicator of multicollinearity. Despite non-normal data distribution, the mixed-effects model is justifiable given our large sample size and satisfied assumptions. Data transformations and processing were performed to further ensure assumption adherence and enhance result reliability.

8.1 Passively Sensed Behavioral Patterns

The results for persistence of passively sensed behavioral patterns are listed in Table 7. Our findings show that the time spent walking decreased by approximately 52 minutes during the COVID period compared to the pre-pandemic level. However, after returning to campus, it increased by almost 3 minutes. The significant decrease in time spent walking during the COVID period compared to the pre-COVID period suggests that the pandemic had a negative impact on walking behavior. However, the increase in time spent walking after returning to campus compared to the pre-COVID period indicates that walking behavior is recovering or returning towards its pre-COVID baseline. Students were more stationary during the COVID period, with an increase in still duration of 62.81 minutes, and continued to have a higher still duration of about 9 minutes after they returned to campus.

Table 7. Behavioral Persistence: The table shows results from a mixed-effects model analyzing behavioral changes during COVID-19 and the transition back to in-person instruction compared to pre-pandemic times. Significance levels: (*** $p < .01$, ** $.01 \leq p < .05$, * $.05 \leq p \leq .10$). The 'Pre-COVID' period is the model's intercept. Columns 'β', 'S.E.', and 'Z' display the coefficient, standard error, and Z statistics (which is a measure of how far a particular data point is from the mean of a dataset, expressed in standard deviation units) for each period, respectively, using only 24-hour average data (epochs 0). Note: Durations are in minutes, except sleep (hours), distances (miles), and sleep start/end times (24-hour format). Coefficients for changes in sleep start and end times during COVID-19 and upon returning to in-person instruction are in minutes.

Behavior/Self-report	Pre-COVID (intercept)			During COVID			In-person instruction		
	β	S.E.	Z	β	S.E.	Z	β	S.E.	Z
Duration in vehicle	29.00	2.41	12.00***	-1.64	0.33	-4.90***	3.00	0.60	-4.96***
Still duration	1239.81	8.70	142.45***	62.81	0.61	102.80***	8.99	1.10	8.16***
Number of steps	7384.70	426.88	17.29***	-3374.32	24.27	-139.00***	-1184.81	43.80	-27.04***
Walking duration	133.18	9.92	13.43***	-51.990	0.43	-119.86***	2.88	0.78	3.67***
Distance travelled	116.06	19.44	5.97***	-38.79	11.86	-3.27***	-40.45	21.41	-1.89*
Sleep duration	7.33	0.19	37.03***	0.17	0.01	13.79***	0.20	0.02	8.89***
Sleep end	07:30	1.26	79.42***	4.00	0.09	5.78***	17.72	0.17	14.29***
Sleep start	00:01	1.38	30.35***	-6.08	0.08	-9.57***	6.00	0.15	5.21***
Unlock duration	181.00	11.18	16.19***	33.85	0.44	76.83***	2.53	0.80	3.18***
Number of unlocks	87.97	6.69	13.14***	-17.06	0.21	-82.01***	-13.24	0.37	-35.26***
Time spent at home	8.62	0.56	15.33***	0.51	0.04	13.57***	-3.22	0.07	-47.08***
Time spent at workout places	0.20	0.05	4.20***	-0.20	0.00	-51.61***	0.11	0.01	15.35***
Audio play duration	110.75	17.72	6.25***	46.40	0.65	71.27***	37.75	1.18	32.14***
Number of audio plays	4.09	0.44	9.30***	1.21	0.01	84.5***	0.83	0.03	32.25***
Time spent at study places	2.62	0.17	15.06***	-2.42	0.01	-155.86***	-1.46	0.03	-56.37***
Number of locations visisted	3.74	0.14	27.28***	-2.64	0.01	-289.92***	-1.97	0.02	-120.11***
Time spent at social places	0.37	0.04	9.17***	-0.32	0.00	-75.14	-0.10	0.01	-13.31
Max distance from campus	1029.44	98.10	10.49***	298.88	5.92	50.43***	-222.85	10.69	-20.84***
Depression	1.03	0.17	6.09***	0.43	0.01	84.72***	0.33	0.01	36.30***
Anxiety	1.40	0.16	8.70***	0.32	0.01	62.77***	0.24	0.01	25.69***
PHQ-4	2.42	0.31	7.68***	0.75	0.01	83.18***	0.57	0.02	35.01***
Self-esteem	7.85	0.18	42.96***	0.22	0.01	34.07***	0.32	0.01	27.03***
Stress	2.58	0.09	26.83***	0.26	0.01	57.18***	0.23	0.01	27.67***
Social level	3.14	0.07	41.86***	-0.33	0.01	-59.07***	-0.28	0.01	-27.74***
COVID concern	-	-	-	4.08	0.21	19.01***	-0.55	0.02	-22.47***

Similarly, students traveled less during the COVID period, with a decrease of 38.78 miles (ca. 62 km), and they continued to travel less after returning to campus (with an even higher decrease of 40.45 miles (ca. 65 km) compared to the during-COVID period, holding all other variables constant). Time spent on a vehicle decreased by 1.64 minutes during the COVID period compared to the pre-COVID duration, and it continued to decrease (with a higher rate of 3 minutes) even after the students returned to campus. Interestingly, one of the interviewed students' perceptions seem to contradict our quantitative results, as they said, *"I'm surprised by the vehicle finding, I was spending a lot more time in the car during COVID. I was driving because it was an activity, going to grocery stores more frequently."* But both of their observations in reduction in vehicle travel after returning to campus correspond with our findings, as they mentioned students were on campus and were not driving or traveling much.

Our findings also demonstrate that the pandemic has influenced students' sleep patterns and phone usage. During the COVID period, sleep duration increased by 0.168 hours compared to the pre-COVID period and continued to rise by 0.196 hours during the in-person instruction period, suggesting that students have maintained a longer sleep duration even after returning to campus. Sleep end hours increased by 0.530 hours during the

COVID period, and this trend continued in the later period with an additional increase of 2.363 hours. In contrast, sleep start hours initially decreased by 0.81 hours during the pandemic, but later increased by 0.79 hours in the post-COVID period, indicating a shift towards later bedtimes. Regarding phone usage, phone unlock duration increased by 33.85 minutes during the COVID period compared to the pre-COVID period, but only increased by 2.53 minutes after returning to in-person instruction, suggesting that phone usage is in the verge of returning to its pre-pandemic levels. A possible reason for increase in phone usage during COVID is to attend the remote classes. One of the interviewed students mentioned, *"there was definitely a huge increase in the amount that I used my phone during COVID because I would be calling my friends for hours and I was just on Zoom or on Facetime with friends."* One of the students recalled that she would also frequently watch movies with her friends, *"Netflix has an extension where you can watch a show with someone else from a different location. So once students were in their room, they probably were on their phones more. I would do that to socialize through phone."* The number of phone unlocks decreased by 17.06 units during the COVID period, and although this number increased slightly during the in-person phase, it still remained lower than the pre-COVID baseline with a decrease of 13.24 unlocks. Audio playing duration increased by 46.40 minutes during the COVID period, and this trend continued after return to in-person instruction, with an increase of 37.76 minutes compared to the pre-COVID level, indicating that students engaged more with audio content during the pandemic as well as after return to campus. The number of audio plays increased during the COVID period by 1.209 units and continued to increase in the latter period by 0.833 units, which supports the observation of increased engagement with audio content. The number of unique locations visited by students decreased by 2.641 during the COVID period, reflecting the impact of lockdowns and restricted movement. However, after returning to campus, the decrease was smaller at 1.975 unique places visited, suggesting that students started to visit more locations as restrictions eased. The maximum distance traveled from campus decreased by 222.85 miles (ca. 358 km) after returning to campus compared to the pre-COVID period, which could be attributed to changes in students' travel patterns or preferences after the pandemic subsides.

One of the students we interviewed found the reduction in steps even after return to in-person instruction interesting. She shared that it might be because a lot of classes had Zoom alternatives even though most of the students were back on campus – *"Every class that was in person was still required to be recorded. So you could not be penalized for missing class. Whereas right now, if you are not there for attendance, you get marked off. Even though you were encouraged to go in person, the professors were also very lenient, encouraging kids to stay home if they did not feel well. I think classes in 2021 and the entire of 2022 were all recorded. So I know people sometimes will skip classes and be like I'll watch the lecture later. A lot of kids had gotten used to not going to classes, so they were either saying I'm not feeling well to watch the recorded lecture or they did actually have COVID and I think that might have reduced the steps."* Another contributing factor, as suggested by a student, could be the growing number of personal vehicles among students in their later college years. This shift results in greater accessibility to rides among peers, reducing the need to depend on upperclassmen for transportation and consequently, possibly leading to a decrease in steps.

8.2 Self-reported Mental Health Outcomes

Our analysis of mental health outcomes reveal important changes in students' mental health during the COVID period and after return to in-person instruction compared to the pre-COVID period. Persistence of self-reported outcomes is also listed in Table 7 towards the bottom. As shown in the table, depression scores increased by 0.430 during the COVID period and continued to be elevated by 0.334 after that. Anxiety scores also increased during the COVID period by 0.322, with a slightly smaller increase of 0.239 after returning to campus. These findings may suggest that the pandemic had a negative impact on students' mental health, with higher levels of depression and anxiety observed during and after the COVID period. The PHQ-4 scores, a measure of psychological distress, increased by 0.750 during the COVID period, indicating heightened distress among students. In the in-person period, the increase was smaller at 0.574, suggesting some improvement in psychological distress after the return to in-person instruction. Self-esteem scores increased marginally by 0.225 during the COVID period and continued to increase by 0.325 after that, indicating a slight improvement in students' self-esteem throughout the pandemic and afterward. One of the interviewed students shared that they believed self-esteem would naturally improve as one got older. But the other mentioned that there could be two sides to it. As one grows older, it could be isolating as groups form and one may either feel a lot better about oneself or a lot worse, depending on whether they are finding their place. Both the students shed light on the impact of COVID in shaping their self-image, with one saying, *"I think something that COVID did especially here was stratified friends and friend groups. If you did not have a really strong friend group since before the pandemic, then it was really hard for a lot of people to talk to others and feel like a friend group and that they belonged to the same college. People who did not have such groups would see all of their friends hang out without them and then when they came for in person, the groups who hung out during COVID would be so much closer because they were around. So they did not feel like they maybe fit into their groups as much."*. The other student supported this remark and mentioned, *"There was a lot more cliques, friend groups were a lot smaller, and that probably either helped improve the self image or decrease it."*

Following that, stress scores increased by 0.260 during the COVID period and remained elevated, with an increase of 0.229 after the period. This suggests that students continued to experience higher levels of stress even after returning to campus. Social levels decreased by 0.326 during the COVID period, reflecting the impact of social distancing and lockdown measures on students' social lives. Although social levels improved slightly after returning to in-person instruction, they remained lower than the pre-COVID baseline, with a decrease of 0.278. This indicates that students' social lives have not yet fully recovered from the effects of the pandemic. COVID concern scores were only available for the during-COVID and in-person instruction periods. During the COVID period, the COVID concern score was 4.076, which decreased by 0.553 during the in-person instruction period. This suggests that students' concerns about COVID-19 have diminished as the pandemic situation improved and they returned to campus. One of the students shared that, *"I think the mental health issues will eventually baseline out a bit more, but it might take a longer time. Because I feel there is a subset of people who realized that they might have issues with depression and anxiety and their mental health that they did not really realize until they were locked in their home alone. And so I think that might eventually go back once they have more social interaction and other care to reduce those levels, but probably not super fast."*

In summary, our findings suggest that the COVID-19 pandemic might had a significant impact on students' behaviors and mental health outcomes. Although we do not have a control group to directly compare changes over time, the unique circumstances and timing of the pandemic provide strong indications of a potential association. Some behaviors, such as time spent walking and phone usage, improved or returned to pre-COVID levels after resuming in-person instruction, while others, like time spent in vehicles, sedentary duration, and distance traveled, continued to show persistent changes. Mental health outcomes also appear to have been affected, with increased levels of depression, anxiety, stress, and psychological distress observed. While self-esteem improved during the decline of the pandemic and in-person instruction period, social engagement levels have not fully recovered. The decrease in COVID concern scores after returning to in-person instruction suggests that students' worries about the pandemic have diminished as they adjusted to the new normal. Overall, the observed changes in students' behaviors and mental health outcomes during the pandemic and after in-person instruction underscore the potential complex and multifaceted impact of COVID-19 on the college population. Monitoring these trends and providing appropriate support remains important as students continue to adapt and return to in-person instruction.

9 DISCUSSION

In this section, we discuss our findings, the implications of our work and the associated ethical considerations.

9.1 Summary of findings

Our data highlights the changes in daily routines and physical activity levels among students throughout their four years at Dartmouth. Particularly noteworthy was the comparison between first and final term students across different cohorts. Notably, the 2021 cohort displayed heightened sedentary behavior, which could be indicative of an increased workload and the postponement of challenging courses or thesis work. In contrast, the 2022 cohort demonstrated greater physical activity and less time spent at home, potentially reflecting the reintroduction of social events and a more vibrant campus life. We also observed gender differences, with male students generally being more active while female students exhibited higher levels of sedentary behavior and expressed more concern about COVID-19. This highlights the possibility of differential health risks and underscores the need for gender-specific interventions [44, 70]. Furthermore, we examined the contrasting experiences of students living on-campus versus those living off-campus during the pandemic. Our findings showed significant behavioral variations between these groups, reinforcing the importance of tailored support and resources to cater to these differing circumstances [7].

The ramifications of the COVID-19 pandemic on student behaviors, as discussed in our study, extend beyond the immediate impact of the crisis. As Dartmouth College pivoted back to in-person instruction, our research observed a persistence of certain behavioral patterns that were adopted during the remote learning period. This phenomenon was particularly evident in physical activity and technology usage. For instance, we noted a sustained decrease in distance travelled and vehicle use even after the resumption of in-person classes. Furthermore, the increased level of stillness during the pandemic period seemed to persist, indicating a possible long-term shift towards more sedentary behaviors. Another interesting observation was related to phone usage. While it was expected that phone usage would increase during the height of the pandemic due to remote learning and reduced face-to-face interactions, we discovered that even with the return to in-person instruction, phone usage did not revert entirely to pre-pandemic levels. This could reflect a lasting shift in communication and study habits, where students continue to rely heavily on digital tools despite the availability of in-person options.

These persistent behavioral changes underscore the profound and enduring impact of the pandemic on students' lifestyles and habits. It also illustrates the adaptability of the student population in embracing new routines and practices in response to external circumstances [37]. Moreover, these observations raise important considerations for universities and policymakers in terms of providing support for these behavioral adjustments, such as enhancing infrastructure for active transportation and providing resources to manage screen time effectively. This persistence of behavioral change is also of significant interest to the HCI field. It underscores the need for adaptive technologies that can seamlessly transition between different contexts (in-person, remote, hybrid) and continue to provide support as users' behaviors evolve [16]. Future HCI research could explore ways to design and implement such flexible technologies, further contributing to the resilience and well-being of students in a post-pandemic world.

9.2 Implications

Our findings have important implications for universities, policymakers, and mental health practitioners, as they reveal key insights into the factors that contribute to students' ability to adapt and thrive during challenging circumstances. By understanding the unique experiences and challenges faced by students during their college journey, we can better support their mental health and well-being, ensuring that they have the resources and tools they need to succeed academically and personally.

For the HCI field, our research highlights the potential of mobile sensing technology and innovative data collection methods in capturing fine-grained, real-time information on student behaviors and mental health, importantly, over long periods of time. This opens up new avenues for the development of context-aware applications and interventions that can support students in managing their well-being throughout their college years. Our study promotes interdisciplinary collaboration between the HCI field and other domains, such as psychology, education, and public health. By working together, experts from various disciplines can develop a more holistic understanding of the complex factors influencing college students' mental health and well-being, leading to more effective and contextually appropriate solutions. Educational institutions can benefit from our findings by implementing targeted support programs and policies that consider the unique challenges and needs of students during different stages of their college journey, including during times of crisis. This could involve fostering a supportive campus environment, offering mental health resources, and creating flexible academic structures that accommodate students' diverse needs. For mental health professionals, our study provides evidence of the associations between daily activities, living arrangements, and mental health outcomes, emphasizing the importance of considering students' behavioral patterns and environmental factors in the development of effective interventions and support strategies. This knowledge can inform the design of tailored, evidence-based approaches that promote resilience and well-being among college students. Many universities are augmenting student heath services with a number of digital tools to support the mental health of students, such as, mindfulness apps and online access to mental health therapists. With the advance of AI and other digital tools there is likely to be a new wave of intervention technologies that could also integrate behavior health driven by mobile sensing. Lastly, our findings have implications for policymakers and public health officials, who can leverage this information to advocate for mental health initiatives and support services targeting college students at a national level. By understanding the broader impact of the pandemic and other environmental factors on students' mental health, policymakers can develop informed strategies and allocate resources to ensure that the well-being of this population remains a priority.

The findings from our comprehensive four-year study can offer valuable insights to student advocacy groups, including Active Minds [46] and the Jed Foundation [22]. Moreover, numerous colleges are actively sharing their annual institutional mental health data, contributing to the national Healthy Minds Study [55]. The inclusion of validated measures from the Healthy Minds Study survey provides essential national data on the prevalence of mental health outcomes and attitudes about mental health on college campuses. Examining mental health across diverse campuses, both locally and nationally, is crucial for ensuring the reproducibility and generalizability of results, as demonstrated in our paper. A collective effort involving additional studies at various types of educational institutions, both at the undergraduate and graduate levels, national survey outcomes, and advocacy by diverse groups will significantly contribute to advancing our understanding of student mental health. This collaborative approach aims to reduce stigma, enhance knowledge, and ultimately eliminate barriers to mental health service utilization, thereby addressing the mental health crisis prevalent on college campuses.

In summary, this study offers valuable insights and implications for various stakeholders involved in the college experience, highlighting the potential of interdisciplinary research and the importance of collaboration between the HCI field, educational institutions, mental health professionals, advocacy groups and policymakers to promote the mental health of college students.

9.3 Ethical Considerations

The design and execution of our study were underpinned by a firm commitment to ethical research practices. All participants were informed about the nature of the study, the type of data collected, and its intended use. Informed consent was obtained from all participants prior to the start of the study. The data collected was anonymized and securely stored, ensuring that individual participants could not be identified. Moreover, privacy was a paramount concern throughout the study, and we took steps to ensure that the data collected did not infringe upon the participants' personal lives. We also respected participants' right to withdraw from the study at any time without penalty. Furthermore, this study was approved by the Institutional Review Board (IRB) of the involved institution, affirming its adherence to ethical research standards. It is important to note that the insights and conclusions drawn from this study are meant to contribute to the broader understanding of college experience, student's behavior change due to COVID and student resilience and are not intended to label or stigmatize individuals or groups.

10 LIMITATIONS AND FUTURE WORK

Despite the valuable insights provided by this study, it is important to acknowledge its limitations, which may have implications for the interpretation and generalizability of the findings.

First, the sample size and demographic of the participants are limited to students from a single institution, which may not be representative of the broader college student population. Next, our analysis is based on self-reported mental health metrics, which may be susceptible to recall bias and social desirability bias. The use of objective measures, such as clinical interviews or assessments, would help to validate and augment the self-reported data. Another limitation is the potential for attrition bias. As the study spans four years, there may be differences between the students who remained in the study and those who dropped out or had incomplete data. These differences could lead to biased estimates of the effects under investigation. This study predominantly focuses on the associations between various behavioral factors and mental health outcomes. While these associations provide valuable insights, they do not necessarily imply causality. Unmeasured confounding factors or other contextual influences may be responsible for the observed associations, and future research should aim to explore the potential causal pathways underlying these relationships. This is also true for the comparison we did between on-campus and off-campus students. Although we used propensity score matching (PSM) to control for potential confounding variables when comparing on-campus and off-campus students, we may not have accounted for all confounding factors. There might be other unobserved variables that could influence the differences observed between the two groups. Similarly, although we have conducted analyses to assess the presence of MCAR and MAR in our dataset, it is important to acknowledge that these analyses might not be entirely reliable due to the complex nature of missing data mechanisms. The methods we employed to detect MCAR and MAR, such as Little's MCAR test and correlation analysis, have their own limitations and assumptions that may not hold in all situations. As a result, our conclusions regarding the presence of MCAR or MAR should be interpreted with caution. Furthermore, it is difficult to directly test for missing not at random (MNAR), as it depends on unobserved data. We acknowledge that our study might be subject to potential biases arising from MNAR, and our conclusions should be interpreted with caution. Given the limitations of our dataset, we cannot definitively rule out the possibility of MNAR. To account for the potential impact of MNAR, as well as the limitations of our MCAR and MAR analyses in future research, it would be beneficial to incorporate additional data sources or use data collection methods that minimize the likelihood of missing data related to unobserved variables. For example, researchers could use wearable devices or smartwatches in combination with smartphones to collect more comprehensive data on participants' behaviors, or they could employ more frequent follow-ups and reminders to encourage participants to provide complete data. Related to this, there are potential biases that may arise due to differences in how participants carry their phones, which could impact our conclusions, such as the observed difference in sedentary behavior between male and female participants. The limitations in our data collection process, including the inability to directly account for how and where participants carry their phones, are important factors to consider when interpreting our findings. Incorporating additional data sources, such as activity trackers or smartwatches could help to obtain a more accurate and comprehensive picture of participants' behaviors and better account for potential biases related to phone usage and carrying habits.

Although our sample is limited to a single institution and may not represent the broader college student population, the findings still offer valuable insights into the experiences of students facing similar challenges during the pandemic. These insights can serve as a foundation for future research and contribute to the design of interventions promoting resilience and well-being among college students in various settings, including low and middle-income countries (LMICs). We recognize the potential limitations in generalizing our findings to LMICs, where students may encounter distinct challenges and access different resources. Nevertheless, the insights obtained from this study can benefit LMICs by providing an understanding of the associations between daily activities, living arrangements, and mental health outcomes during a global crisis. This knowledge can guide the design of interventions and support strategies tailored to the unique needs and circumstances of students in LMICs. To further enhance the generalizability of our findings, we recommend future research involve multi-institutional collaborations across various countries, including LMICs. This approach would enable researchers to capture a more diverse range of student experiences, resulting in a comprehensive understanding of the factors influencing college students' mental health and well-being during challenging times. In addition, incorporating qualitative data by interviewing a larger and more diverse group of students would enrich the insights provided by this study. In this study, we employed multiple single-item measures, such as stress, social level, and COVID concern, which may not fully capture the complexity of these constructs. Our choice to use single-item measures was motivated by our desire to reduce participant burden and encourage higher response rates, given the repeated nature of our data collection. We were able to find some evidence of criterion validity post-study by correlating these single-item scores with other mental health outcomes. For example, we examined the association between the single-item stress measure and the PHQ4, finding a strong positive correlation between stress and the total PHQ4 score (ρ=0.64), depression (ρ=0.58), and anxiety (ρ=0.63) subcomponents, which are known to be associated with stress. But still, we acknowledge that more comprehensive measures, such as validated multi-item scales, could provide a more accurate representation of students' stress, social level, and COVID concern throughout the semester. Future research should consider using more extensive measures to assess these constructs and further validate the findings of this study.

11 COLLEGE EXPERIENCE STUDY DATASET AND OPEN RESEARCH QUESTIONS

In the interest of advancing research into the mental health of college students during their college years using mobile sensing and self-reports, we discuss the public release of the de-identified College Experience Study dataset on Kaggle [26]. We also outline a series of open research questions that could be studied using the public dataset. We encourage the use of this dataset for academic and research purposes and request that any publications resulting from its use appropriately cite it as follows:

> Subigya Nepal, Wenjun Liu, Arvind Pillai, Weichen Wang, Vlado Vojdanovski, Jeremy F. Huckins, Courtney Rogers, Meghan L. Meyer, and Andrew T. Campbell. 2024. Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience and Behavior of College Students during the Pandemic. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* 8, 1, Article 39 (March 2024), 37 pages.
>
> **Dataset URL: https://www.kaggle.com/datasets/subigyanepal/college-experience-dataset**

We hope that this contribution will facilitate a deeper understanding of the pandemic's impact on student life or the student life in general and promote further studies in this domain. There are many open questions that researchers can study with this dataset; for example

- **Student Behavior During National Events:** The longitudinal data spanning four years offers a unique opportunity to analyze student behavioral shifts in response to national events, such as electoral cycles.

Research could investigate how significant societal moments correlate with changes in student mood, activity levels, and social interactions.

- **Cross-Institutional Comparisons:** The unique dataset allows for broader comparisons, leading to the question: How do the behavioral and mental health trends observed in this study compare with those at other institutions, both within the US and internationally? Our dataset can act as a benchmark for educational institutions to identify pervasive behavioral patterns and institution-specific trends, aiding in the creation of targeted mental health initiatives.
- **Model Pre-training and Transfer Learning:** Smaller studies often face constraints in data availability. Leveraging our extensive dataset for pre-training models could enhance their predictive performance through transfer learning, thereby optimizing research efficacy and reducing resource expenditure.
- **Longitudinal Mental Health Trajectories:** How do mental health trajectories evolve over the course of college life? The dataset allows for an examination of long-term trends and the identification of critical periods for intervention.
- **Socioeconomic Status and Social Dynamics:** Investigating the influence of self-perceived socioeconomic status on students' social interactions might provide insights into social stratification within college environments.

12 CONCLUSION

In conclusion, our longitudinal mobile sensing study provides a comprehensive analysis of undergraduate students' behaviors, mental health metrics, and experiences throughout their four-year college journey at Dartmouth College, with a particular focus on the unprecedented challenges presented by the COVID-19 pandemic. Through the utilization of a mobile sensing app and supplementary qualitative insights from student interviews, we identified patterns and associations between daily activities, living arrangements, and mental health outcomes. This study not only offers valuable insights into the impact of the pandemic on college students but also emphasizes the significance of understanding the broader context of their experiences throughout their college years. The findings of this study contribute to the existing body of literature on the relationship between daily activities and mental health, highlighting the ongoing need for research and support systems that foster resilience and well-being among college students. The study also underscores the potential of mobile sensing technology and interdisciplinary approaches to illuminate the interplay between behavior, environment, and mental health within higher education settings. Overall, this study emphasizes the importance of a holistic understanding of college students' experiences and provides a foundation for the development of targeted interventions and support mechanisms that address the unique challenges they face. Finally, we publicly released the College Experience dataset [26] to help accelerate research in this field.

ACKNOWLEDGMENTS

We extend our deepest gratitude to the National Institute of Mental Health for their generous support through grant number 5R01MH059282, which was instrumental in the realization of this extensive four-year study. Our heartfelt thanks go out to the students who volunteered their time and effort, playing a pivotal role in the success of our research. We also wish to acknowledge the collective effort and unwavering support of everyone involved in this project. This research endeavor has been greatly enriched by the collaborative spirit and shared vision of all participants, making it not just a study, but a journey of discovery and learning. To each and every one of you, thank you for being an essential part of this meaningful undertaking.

REFERENCES

[1] Saeed Abdullah, Elizabeth L. Murnane, Mark Matthews, Matthew Kay, Julie A. Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive Rhythms: Unobtrusive and Continuous Sensing of Alertness Using a Mobile Phone. In *Proceedings of the 2016 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 178–189. https://doi.org/10.1145/2971648.2971712

[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] Daniel A Adler, Vincent W-S Tseng, Gengmo Qi, Joseph Scarpa, Srijan Sen, and Tanzeem Choudhury. 2021. Identifying mobile sensing indicators of stress-resilience. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5, 2 (2021), 1–32.

[4] Daniel A Adler, Fei Wang, David C Mohr, and Tanzeem Choudhury. 2022. Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies. *Plos one* 17, 4 (2022), e0266516.

[5] Muhammad Adnan and Kainat Anwar. 2020. Online Learning amid the COVID-19 Pandemic: Students' Perspectives. *Online Submission* 2, 1 (2020), 45–51.

[6] WP Anderson Jr and Sandra I Lopez-Baez. 2012. College-Student Personal-Growth and Attributions of Cause. *Professional Counselor* 2, 1 (2012), 66–80.

[7] Anthony Lising Antonio. 2017. Student Life as Relational Networks. Connecting in College: How Friendship Networks Matter for Academic and Social Success by Janice M. McCabe. *Journal of College Student Development* 58, 3 (2017), 472–474. https://doi.org/10.1353/csd.2017.0037

[8] Callum R. K. Arnold, Sreenidhi Srinivasan, Sophie Rodriguez, Natalie Rydzak, Catherine M. Herzog, Abhinay Gontu, Nita Bharti, Meg Small, Connie J. Rogers, Margeaux M. Schade, Suresh V. Kuchipudi, Vivek Kapur, Andrew F. Read, and Matthew J. Ferrari. 2022. A longitudinal study of the impact of university student return to campus on the SARS-CoV-2 seroprevalence among the community members. *Scientific Reports* 12, 1 (May 2022). https://doi.org/10.1038/s41598-022-12499-5

[9] Jakob E. Bardram and Aleksandar Matic. 2020. A Decade of Ubiquitous Computing Research in Mental Health. *IEEE Pervasive Computing* 19, 1 (Jan. 2020), 62–72. https://doi.org/10.1109/mprv.2019.2925338

[10] Dror Ben-Zeev, Rachel Brian, Rui Wang, Weichen Wang, Andrew T. Campbell, Min S. H. Aung, Michael Merrill, Vincent W. S. Tseng, Tanzeem Choudhury, Marta Hauser, John M. Kane, and Emily A. Scherer. 2017. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatric Rehabilitation Journal* 40, 3 (Sept. 2017), 266–275. https://doi.org/10.1037/prj0000243

[11] Tjeerd W Boonstra, Jennifer Nicholas, Quincy JJ Wong, Frances Shaw, Samuel Townsend, and Helen Christensen. 2018. Using mobile phone sensor technology for mental health research: integrated analysis to identify hidden challenges and potential solutions. *Journal of medical Internet research* 20, 7 (2018), e10131.

[12] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. https://doi.org/10.1145/2750858.2805845

[13] Wenjun Cao, Ziwei Fang, Guoqiang Hou, Mei Han, Xinrong Xu, Jiaxin Dong, and Jianzhong Zheng. 2020. The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research* 287 (May 2020), 112934. https://doi.org/10.1016/j.psychres.2020.112934

[14] 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*. 145–152.

[15] Prerna Chikersal, Afsaneh Doryab, Michael Tumminia, Daniella K. Villalba, Janine M. Dutcher, Xinwen Liu, Sheldon Cohen, Kasey G. Creswell, Jennifer Mankoff, J. David Creswell, Mayank Goel, and Anind K. Dey. 2021. Detecting Depression and Predicting Its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 3 (jan 2021), 41 pages. https://doi.org/10.1145/3422821

[16] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A. Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/1124772.1124840

[17] Mark É. Czeisler, Emily R. Capodilupo, Matthew D. Weaver, Charles A. Czeisler, Mark E. Howard, and Shantha M.W. Rajaratnam. 2022. Prior sleep-wake behaviors are associated with mental health outcomes during the COVID-19 pandemic among adult users of a wearable device in the United States. *Sleep Health* 8, 3 (June 2022), 311–321. https://doi.org/10.1016/j.sleh.2022.03.001

[18] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–27.

[19] Lianne P. de Vries, Bart M. L. Baselmans, and Meike Bartels. 2020. Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. *Journal of Happiness Studies* (Oct. 2020). https://doi.org/10.1007/s10902-020-00324-7

[20] Afsaneh Doryab, Prerna Chikarsel, Xinwen Liu, and Anind K Dey. 2018. Extraction of behavioral features from smartphone and wearable data. *arXiv preprint arXiv:1812.10394* (2018).

[21] Maria Faurholt-Jepsen, Jonas Busk, Helga Þórarinsdóttir, Mads Frost, Jakob Eyvind Bardram, Maj Vinberg, and Lars Vedel Kessing. 2018. Objective smartphone data as a potential diagnostic marker of bipolar disorder. *Australian amp; New Zealand Journal of Psychiatry* 53, 2 (Nov. 2018), 119–128. https://doi.org/10.1177/0004867418808900

[22] The Jed Foundation. 2023. *The Jed Foundation*. https://jedfoundation.org

[23] Gabriella M Harari, Sandrine R Müller, Min SH Aung, and Peter J Rentfrow. 2017. Smartphone sensing methods for studying behavior in everyday life. *Current opinion in behavioral sciences* 18 (2017), 83–90.

[24] The Lancet Public Health. 2019. Time to tackle the physical activity gender gap. *The Lancet Public Health* 4, 8 (Aug. 2019), e360. https://doi.org/10.1016/s2468-2667(19)30135-5

[25] Jeremy F Huckins, Alex W DaSilva, Weichen Wang, Elin Hedlund, Courtney Rogers, Subigya K Nepal, Jialing Wu, Mikio Obuchi, Eilis I Murphy, Meghan L Meyer, et al. 2020. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. *Journal of medical Internet research* 22, 6 (2020), e20185.

[26] Kaggle Inc. 2024. *Public release of College Experience Study Dataset available on Kaggle*. https://www.kaggle.com/datasets/subigyanepal/college-experience-dataset/

[27] Wenning Jiang, Jin Luo, and Hannan Guan. 2021. Gender Difference in the Relationship of Physical Activity and Subjective Happiness Among Chinese University Students. *Frontiers in Psychology* 12 (Dec. 2021). https://doi.org/10.3389/fpsyg.2021.800515

[28] K C Kling, J S Hyde, C J Showers, and B N Buswell. 1999. Gender differences in self-esteem: a meta-analysis. *Psychol. Bull.* 125, 4 (July 1999), 470–500.

[29] Seyma Kucukozer-Cavdar, Tugba Taskaya-Temizel, Abhinav Mehrotra, Mirco Musolesi, and Peter Tino. 2021. Designing Robust Models for Behaviour Prediction Using Sparse Data from Mobile Sensing: A Case Study of Office Workers' Availability for Well-being Interventions. *ACM Transactions on Computing for Healthcare* 2, 4 (2021), 1–33.

[30] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications magazine* 48, 9 (2010), 140–150.

[31] Michael E Lauderdale, Sami Yli-Piipari, Carol C Irwin, and Todd E Layne. 2015. Gender differences regarding motivation for physical activity among college students: A self-determination approach. *The Physical Educator* 72, 5 (2015).

[32] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM. https://doi.org/10.1145/2462456.2464449

[33] Jonathan Liono, Zahraa S. Abdallah, A. K. Qin, and Flora D. Salim. 2018. Inferring Transportation Mode and Human Activity from Mobile Sensing in Daily Life. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (New York, NY, USA) *(MobiQuitous '18)*. Association for Computing Machinery, New York, NY, USA, 342–351. https://doi.org/10.1145/3286978.3287006

[34] Elena Losina, Valia Leifer, Lucia Millham, Christopher Panella, Emily P. Hyle, Amir M. Mohareb, Anne M. Neilan, Andrea L. Ciaranello, Pooyan Kazemian, and Kenneth A. Freedberg. 2021. College Campuses and COVID-19 Mitigation: Clinical and Economic Value. *Annals of Internal Medicine* 174, 4 (April 2021), 472–483. https://doi.org/10.7326/m20-6558

[35] Lucy MacLeod, Banuchitra Suruliraj, Dominik Gall, Kitti Bessenyei, Sara Hamm, Isaac Romkey, Alexa Bagnell, Manuel Mattheisen, Viswanath Muthukumaraswamy, Rita Orji, and Sandra Meier. 2021. A Mobile Sensing App to Monitor Youth Mental Health: Observational Pilot Study. *JMIR mHealth and uHealth* 9, 10 (Oct. 2021), e20638. https://doi.org/10.2196/20638

[36] Marcia B Baxter Magolda. 2006. Intellectual development in the college years. *Change: The Magazine of Higher Learning* 38, 3 (2006), 50–54.

[37] Andrew J. Martin, Paul Ginns, and Rebecca J. Collie. 2023. University students in COVID-19 lockdown: The role of adaptability and fluid reasoning in supporting their academic motivation and engagement. *Learning and Instruction* 83 (Feb. 2023), 101712. https://doi.org/10.1016/j.learninstruc.2022.101712

[38] Stijn A A Massar, Alyssa S C Ng, Chun Siong Soon, Ju Lynn Ong, Xin Yu Chua, Nicholas I Y N Chee, Tih Shih Lee, and Michael W L Chee. 2021. Reopening after lockdown: the influence of working-from-home and digital device use on sleep, physical activity, and wellbeing following COVID-19 lockdown and reopening. *Sleep* 45, 1 (Oct. 2021). https://doi.org/10.1093/sleep/zsab250

[39] 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, Ge Gao, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Kizito Masaba, Shayan Mirjafari, Edward Moskal, Raghu Mulukutla, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 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* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299041

[40] Lakmal Meegahapola, William Droz, Peter Kun, Amalia de Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, et al. 2023. Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–32.

[41] Lakmal Meegahapola and Daniel Gatica-Perez. 2020. Smartphone sensing for the well-being of young adults: A review. *IEEE Access* 9 (2020), 3374–3399.

[42] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 83 (sep 2017), 21 pages. https://doi.org/10.1145/3130948

[43] Jennifer Melcher, Joel Lavoie, Ryan Hays, Ryan D'Mello, Natali Rauseo-Ricupero, Erica Camacho, Elena Rodriguez-Villa, Hannah Wisniewski, Sarah Lagan, Aditya Vaidyam, et al. 2021. Digital phenotyping of student mental health during COVID-19: an observational study of 100 college students. *Journal of American College Health* (2021), 1–13.

[44] Francesco Menegoni, Manuela Galli, Elena Tacchini, Luca Vismara, Marco Cavigioli, and Paolo Capodaglio. 2009. Gender-specific Effect of Obesity on Balance. *Obesity* 17, 10 (Oct. 2009), 1951–1956. https://doi.org/10.1038/oby.2009.82

[45] Mike A Merrill and Tim Althoff. 2022. Self-supervised Pretraining and Transfer Learning Enable Flu and COVID-19 Predictions in Small Mobile Sensing Datasets. *arXiv preprint arXiv:2205.13607* (2022).

[46] Active Minds. 2023. *Active Minds.* https://www.activeminds.org

[47] 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.

[48] David C Mohr, Mi Zhang, and Stephen M Schueller. 2017. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology* 13 (2017), 23–47.

[49] Margaret E. Morris, Kevin S. Kuehn, Jennifer Brown, Paula S. Nurius, Han Zhang, Yasaman S. Sefidgar, Xuhai Xu, Eve A. Riskin, Anind K. Dey, Sunny Consolvo, and Jennifer C. Mankoff. 2021. College from home during COVID-19: A mixed-methods study of heterogeneous experiences. *PLOS ONE* 16, 6 (June 2021), e0251580. https://doi.org/10.1371/journal.pone.0251580

[50] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K D'Mello, Munmun De Choudhury, Gregory D Abowd, and Thomas Plötz. 2019. Prediction of mood instability with passive sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–21.

[51] 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

[52] Subigya Nepal, Gonzalo J Martinez, Shayan Mirjafari, Koustuv Saha, Vedant Das Swain, Xuhai Xu, Pino G Audia, Munmun De Choudhury, Anind K Dey, Aaron Striegel, et al. 2022. A Survey of Passive Sensing in the Workplace. *arXiv preprint arXiv:2201.03074* (2022).

[53] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–28.

[54] 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 *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[55] The Healthy Minds Network. 2023. *Healthy Minds Study.* https://healthymindsnetwork.org/hms/

[56] Kana Okano, Jakub R. Kaczmarzyk, Neha Dave, John D. E. Gabrieli, and Jeffrey C. Grossman. 2019. Sleep quality, duration, and consistency are associated with better academic performance in college students. *npj Science of Learning* 4, 1 (Oct. 2019). https://doi.org/10.1038/s41539-019-0055-z

[57] Arvind Pillai, Subigya Nepal, and Andrew Campbell. 2023. Rare Life Event Detection via Mobile Sensing Using Multi-Task Learning. In *Conference on Health, Inference, and Learning*. PMLR, 279–293.

[58] Arvind Pillai, Subigya Kumar Nepal, Weichen Wang, Matthew Nemesure, Michael Heinz, George Price, Damien Lekkas, Amanda C Collins, Tess Griffin, Benjamin Buck, et al. 2024. Investigating Generalizability of Speech-based Suicidal Ideation Detection Using Mobile Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–38.

[59] Arcady A. Putilov, Dmitry S. Sveshnikov, Zarina B. Bakaeva, Elena B. Yakunina, Yuri P. Starshinov, Vladimir I. Torshin, Nikolay N. Alipov, Olga V. Sergeeva, Elena A. Trutneva, Michael M. Lapkin, Zhanna N. Lopatskaya, Roman O. Budkevich, Elena V. Budkevich, Alexandra N. Puchkova, and Vladimir B. Dorokhov. 2021. Differences between male and female university students in sleepiness, weekday sleep loss, and weekend sleep duration. *Journal of Adolescence* 88, 1 (March 2021), 84–96. https://doi.org/10.1016/j.adolescence.2021.02.006

[60] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *Proceedings of the 6th International Conference on Digital Health Conference* (Montréal, Québec, Canada) *(DH '16)*. Association for Computing Machinery, New York, NY, USA, 141–150. https://doi.org/10.1145/2896338.2896359

[61] Paul R. Rosenbaum and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55. https://doi.org/10.1093/biomet/70.1.41

[62] Koustuv Saha, Ted Grover, Stephen M Mattingly, Vedant Das Swain, Pranshu Gupta, Gonzalo J Martinez, Pablo Robles-Granda, Gloria Mark, Aaron Striegel, and Munmun De Choudhury. 2021. Person-centered predictions of psychological constructs with social media

contextualized by multimodal sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–32.

[63] Pradeep Sahu. 2020. Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff. *Cureus* 12, 4 (2020).

[64] Akane Sano, Andrew J Phillips, Z Yu Amy, Andrew W McHill, Sara Taylor, Natasha Jaques, Charles A Czeisler, Elizabeth B Klerman, and Rosalind 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)*. IEEE, 1–6.

[65] Natalia Sauer, Agnieszka Sałek, Wojciech Szlasa, Tomasz Ciecieląg, Jakub Obara, Sara Gaweł, Dominik Marciniak, and Katarzyna Karłowicz-Bodalska. 2022. The Impact of COVID-19 on the Mental Well-Being of College Students. *International Journal of Environmental Research and Public Health* 19, 9 (April 2022), 5089. https://doi.org/10.3390/ijerph19095089

[66] Yasaman S Sefidgar, Woosuk Seo, Kevin S Kuehn, Tim Althoff, Anne Browning, Eve Riskin, Paula S Nurius, Anind K Dey, and Jennifer Mankoff. 2019. Passively-sensed behavioral correlates of discrimination events in college students. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–29.

[67] Changwon Son, Sudeep Hegde, Alec Smith, Xiaomei Wang, and Farzan Sasangohar. 2020. Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. *Journal of Medical Internet Research* 22, 9 (Sept. 2020), e21279. https://doi.org/10.2196/21279

[68] Shaoxiong Sun, Amos A Folarin, Yatharth Ranjan, Zulqarnain Rashid, Pauline Conde, Callum Stewart, Nicholas Cummins, Faith Matcham, Gloria Dalla Costa, Sara Simblett, Letizia Leocani, Femke Lamers, Per Soelberg Sørensen, Mathias Buron, Ana Zabalza, Ana Isabel Guerrero Pérez, Brenda WJH Penninx, Sara Siddi, Josep Maria Haro, Inez Myin-Germeys, Aki Rintala, Til Wykes, Vaibhav A Narayan, Giancarlo Comi, Matthew Hotopf, and Richard JB Dobson and. 2020. Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19. *Journal of Medical Internet Research* 22, 9 (Sept. 2020), e19992. https://doi.org/10.2196/19992

[69] Benjamin Tag, Zhanna Sarsenbayeva, Anna L Cox, Greg Wadley, Jorge Goncalves, and Vassilis Kostakos. 2022. Emotion trajectories in smartphone use: Towards recognizing emotion regulation in-the-wild. *International Journal of Human-Computer Studies* 166 (2022), 102872.

[70] Simone Catharina Maria Wilhelmina Tummers, Arjen Hommersom, Lilian Lechner, Roger Bemelmans, and Catherine Adriana Wilhelmina Bolman. 2022. Determinants of physical activity behaviour change in (online) interventions, and gender-specific differences: a Bayesian network model. *International Journal of Behavioral Nutrition and Physical Activity* 19, 1 (Dec. 2022). https://doi.org/10.1186/s12966-022-01381-2

[71] Baruch Vainshelboim, Gabrielle M. Brennan, Stephen LoRusso, Patricia Fitzgerald, and Kristofer S. Wisniewski. 2019. Sedentary behavior and physiological health determinants in male and female college students. *Physiology &amp Behavior* 204 (May 2019), 277–282. https://doi.org/10.1016/j.physbeh.2019.02.041

[72] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–40.

[73] 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.

[74] 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* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 295–306. https://doi.org/10.1145/2750858.2804251

[75] 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.

[76] Weichen Wang, Subigya Nepal, Jeremy F Huckins, Lessley Hernandez, Vlado Vojdanovski, Dante Mack, Jane Plomp, Arvind Pillai, Mikio Obuchi, Alex Dasilva, et al. 2022. First-gen lens: Assessing mental health of first-generation students across their first year at college using mobile sensing. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 6, 2 (2022), 1–32.

[77] 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*. 425–434.

[78] Zhiyuan Wang, Haoyi Xiong, Mingyue Tang, Mehdi Boukhechba, Tabor E Flickinger, and Laura E Barnes. 2022. Mobile Sensing in the COVID-19 Era: A Review. *Health Data Science* (2022).

[79] Ingrid K. Weigold, Arne Weigold, Shu Ling, and Migyeong Jang. 2020. College as a Growth Opportunity: Assessing Personal Growth Initiative and Self-determination Theory. *Journal of Happiness Studies* 22, 5 (Sept. 2020), 2143–2163. https://doi.org/10.1007/s10902-020-

00312-x

[80] Brenton M Wiernik, Deniz S Ones, Benjamin M Marlin, Casey Giordano, Stephan Dilchert, Brittany K Mercado, Kevin C Stanek, Adib Birkland, Yilei Wang, Brenda Ellis, et al. 2020. Using mobile sensors to study personality dynamics. *European Journal of Psychological Assessment* 36, 6 (2020), 935.

[81] Congyu Wu, Hagen Fritz, Melissa Miller, Cameron Craddock, Kerry Kinney, Darla Castelli, and David Schnyer. 2021. Exploring Post COVID-19 Outbreak Intradaily Mobility Pattern Change in College Students: A GPS-Focused Smartphone Sensing Study. *Frontiers in Digital Health* 3 (Nov. 2021). https://doi.org/10.3389/fdgth.2021.765972

[82] Jiaqi Xiong, Orly Lipsitz, Flora Nasri, Leanna M.W. Lui, Hartej Gill, Lee Phan, David Chen-Li, Michelle Iacobucci, Roger Ho, Amna Majeed, and Roger S. McIntyre. 2020. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders* 277 (Dec. 2020), 55–64. https://doi.org/10.1016/j.jad.2020.08.001

[83] 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.

[84] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S Kuehn, Jeremy F Huckins, Margaret E Morris, et al. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34.

[85] Xuhai Xu, Han Zhang, Yasaman Sefidgar, Yiyi Ren, Xin Liu, Woosuk Seo, Jennifer Brown, Kevin Kuehn, Mike Merrill, Paula Nurius, Shwetak Patel, Tim Althoff, Margaret Morris, Eve Riskin, Jennifer Mankoff, and Anind Dey. 2023. GLOBEM Dataset: Multi-Year Datasets for Longitudinal Human Behavior Modeling Generalization. https://doi.org/10.13026/R9S1-S711

[86] Virgil Zeigler-Hill and E.M. Myers. 2012. A review of gender differences in self-esteem. (01 2012), 131–143.

[87] Han Zhang, Paula Nurius, Yasaman Sefidgar, Margaret Morris, Sreenithi Balasubramanian, Jennifer Brown, Anind K Dey, Kevin Kuehn, Eve Riskin, Xuhai Xu, et al. 2020. How does COVID-19 impact students with disabilities/health concerns? *arXiv preprint arXiv:2005.05438* (2020).