Year-Long Digital Phenotyping and Natural Language Processing of Daily Voice Diaries Reveal Affective and Behavioral Signatures of Real-World Life Stress

Constanza M. Vidal Bustamante[a,b], Garth Coombs III[a,b], Habiballah Rahimi-Eichi[a-d], Patrick Mair[a], Jukka-Pekka Onnela[e], Justin T. Baker[c,d], and Randy L. Buckner[a-c,f]

[a] Department of Psychology, Harvard University, Cambridge, Massachusetts, USA

[b] Center for Brain Science, Harvard University, Cambridge, Massachusetts, USA

[c] Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA

[d] Institute for Technology in Psychiatry, McLean Hospital, Belmont, Massachusetts, USA

[e] Department of Biostatistics, Harvard University, Boston, Massachusetts, USA

[f] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Charlestown, Massachusetts, USA

Correspondence should be addressed to Constanza M. Vidal Bustamante at cvidal@g.harvard.edu

Abstract

Experiencing stressful life events has been associated with poor mental health outcomes, but less is known about daily stressful events' more proximate impacts on daily behavior and affect. Here we leverage mobile and wearable technology and recent advances in natural language processing for a fine-grained examination of first-year college students' daily life stress, with a focus on academic and social domains, over a full academic year (8,000+ total daily observations). Experiences of stress were characterized from participants' daily voice diaries narrating the main events of the day, using a combination of expert human labeling and large language models fine-tuned for sentiment analysis and topic modeling. Bayesian hierarchical models assessed within-person associations between diary-derived instances of academic and social stressful events and same-day sleep, physical activity, social activity, and negative affect measured with actigraphy wristbands and phone surveys. Days with academic stressful events were associated with shorter sleep duration, decreased physical activity, reduced desire to be around others, and modestly increased negative affect. Additionally, days with academic stressful events had significantly reduced social interaction and increased time spent on schoolwork relative to days with social stressful events. Meanwhile, days with social stressful events stood out by especially heightened negative affect, above and beyond the effect of academic stressful events. Our results suggest that academic and social dimensions of life stress may present distinct signatures in daily affect and behavior, with potential implications for long-term wellbeing.

Keywords: Stressful life events; digital phenotyping; natural language processing; sleep; negative affect; mental health

The transition to college typically involves changes in multiple aspects of daily life. In addition to frequent class meetings and assessments, students must adapt to an unfamiliar social environment, develop new friendships, and manage their health and finances with increasing independence (Ross, Biebling and Heckert, 1999; Gerdes & Mallinckrodt, 1994). Although these are normative demands associated with college life, they can nevertheless contribute to the substantial rates of psychological distress and mental health disorders in this population (Beiter et al., 2015; Dyson & Renk, 2006; Ross et al., 1999; Thorley, 2017), with about one in three college students reporting experiencing serious psychological distress and a diagnosis of anxiety and/or depression (American College Health Association, 2023; Lipson et al., 2022).

The impact of daily life stress on long-term wellbeing may vary depending on the domain of life stress. Although academic stressors are more common among students (American College Health Association, 2023), interpersonal stressors in daily life have been linked to greater rises in negative mood than other types of stressors like overload at work or the household (Almeida, 2005; Bolger et al., 1989). Relatedly, we recently reported that perceived interpersonal stress, although less frequent than perceived academic stress, was a more sensitive marker of students' longer-term psychological distress (Vidal Bustamante et al., 2022). More broadly, interpersonal stressors, such as family and peer conflict, have been presented as particularly important markers and antecedents of internalizing psychopathology, including among children, adolescents, and adults (Brown & Harris, 1978; O'Neill et al., 2004; Parker et al., 2015; Rudolph et al., 2009, 2016; Whisman & Beach, 2010).

The effects of life stress on long-term wellbeing may be influenced by its more proximate impacts on affective states and behavioral routines (Smyth et al., 2018, 2013). Experiencing stressful events has been broadly associated with rises in negative affect (Bolger et al., 1989; Lazarus, 1999; Zautra, 2006) as well as reduced or irregular sleep duration (Kim & Dimsdale, 2007; Sadeh & Gruber, 2002; Van Reeth et al., 2000; Vidal Bustamante et al., 2020) and decreases in physical (Stults-Kolehmainen & Sinha, 2014; Tomiyama, 2019) and social activity (Cacioppo & Hawkley, 2003; daSilva et al., 2021), all of which have also been linked to poor mental health outcomes (Charles et al., 2013; Cohen, 2004; Teychenne et al., 2015; Zhang et al., 2017). Critically, the pattern and/or magnitude of the effects of daily life stress on affect and behavior may differ between academic and social domains of stress, which in turn may shape distinct longer-term impacts on wellbeing. However, methodological limitations have impeded a

fine-grained, multidimensional examination of life stress. Diary methods have enabled the study of daily stressful events, but these studies have been typically short in duration, collapsed across different domains of life stress, and/or collected measures of negative affect but not other aspects of daily behavior (Almeida, 2005; Charles et al., 2013).

Recent advances in digital technology and natural language processing are enabling the characterization of daily life with unprecedent detail. Pen-and-paper diaries and checklists formerly used to record participants' daily life and stressful events (Stone et al., 1998; Tennen et al., 2000; Bolger et al., 1989; Bolger & Eckenrode, 1991) can now be collected more effectively, efficiently, and at scale via smartphones (Smyth et al., 2023; Vidal Bustamante et al., 2022), and recent advances in artificial intelligence applied to natural language processing (Grootendorst, 2022; Rosenthal et al., 2019) can help automate the analysis and classification of daily diaries' content. Additionally, wearable devices' sensing streams like accelerometer and GPS can seamlessly track sleep, physical activity, and mobility patterns as participants go through their daily lives (Harari et al., 2016; Sano et al., 2018; Vidal Bustamante et al., 2024; Wang et al., 2014). Together, these tools are opening new opportunities for the study of life stress, its impacts on daily affect and behavior, and how these dynamics go on to shape long-term health and wellbeing.

In this paper, we leverage mobile and wearable technology and cutting-edge natural language processing techniques to conduct a deep phenotyping study of 49 first-year college students as they were naturally exposed to academic and social stressful events over a full academic year (10,000+ total daily observations). Our aims were twofold. First, we used daily voice diaries and a combination of large language models and expert human labeling to characterize students' experiences of stress across the year, as well as to identify the specific days when participants experienced academic and/or social stressful events. Second, we leveraged wearable- and smartphone-based metrics of students' sleep duration, physical activity, and self-reported affect, schoolwork, and social activities to examine whether days with academic stressful events and days with social stressful events were associated with changes in affect and behavior within individuals. We hypothesized that days with social stressful events would be associated with heightened levels of negative affect and less time spent socializing. Moreover, we expected that days with academic stressful events would be associated with lower

levels of negative affect relative to days with social events, shorter sleep duration, reduced physical activity, and more time spent on schoolwork.

Method

Participants

Forty-nine undergraduates beginning their first year of college volunteered for a year-long study (46 were 18 years old and three 19 years old; 25 female and 24 male). Portions of the survey and actigraphy data from these participants have been previously reported by the authors in (Vidal Bustamante et al., 2022, 2024), but the voice diary data has not been previously reported. Relevant methods reported previously are repeated here. Students living on campus were recruited via flyers posted on campus boards and distributed via email lists and were enrolled in the pilot study during the first two weeks of their fall semester. Participants were required to be taking full-time classes and own a smartphone compatible with the study smartphone application, Beiwe, which is part of the open-source Beiwe platform for digital phenotyping (Onnela et al., 2021). Students were not excluded for current or past psychiatric disorders or medication use, nor if they began treatment or medication for mental health issues during the study. Interested participants scheduled an in-person consent session where study procedures were explained. Initially 68 participants enrolled in the year-long study, with 19 excluded based on issues with data acquisition: early withdrawal from the study (n=7), technical failure to acquire actigraphy data (n=1), poor quality actigraphy data (n=2), and completion of less than 100 of the daily surveys (n=9). Of the final sample (N=49), 63% identified as White, 14% Black, 10% Asian, 4% American Indian, and 4% Mixed-Race. Twelve percent reported prior diagnosis of a psychiatric disorder (including anxiety, depression, and/or ADHD); 8% maintained active diagnoses. 94% of the participants were iPhone users and 6% were Android users.

Study Design

Informed consent was obtained from all participants and all study procedures and methods were approved by the Institutional Review Board of Harvard University. This intensive longitudinal observational study collected data over a full academic year extending a few days

into the summer break. As originally described in (Vidal Bustamante et al., 2022), this study sought to acquire passive and active data as participants went through their daily lives. Participants completed smartphone-based daily surveys and a voice-recorded diary, wore an actigraphy wristband for continuous activity and sleep monitoring for the duration of the study, completed a battery of online questionnaires at the beginning, middle, and end of the study, and attended brief in-person check-ins every 3-4 weeks. Participants were compensated $1 per each daily survey and $1 for each daily voice diary they submitted, $1 per day for continuously wearing the actigraphy wristband, and $20 per hour for online surveys and attending in-person visits. Milestone bonus payments for completing half of the study ($100) and the full study ($300) were also provided to reward participants' continued compliance.

Measures and Quality Control

Daily Phone-Based Surveys

Smartphone surveys were administered via the Beiwe application (Onnela et al., 2021). Each night before bed participants completed a 46-item self-report survey related to their daily lives. Questions assessed a broad range of behaviors and internal states over the past 24 hours, including sleep quality, stress levels and sources, positive and negative affect, general physical health, daily consumption habits, studying behaviors, and sociability and support (Vidal Bustamante et al., 2022). Most questions were answered using a 5-point Likert scale. Here we report analyses of daily survey questions that assessed time spent on schoolwork, time spent interacting with others, desire to spend time with others, and negative affect. The full wording of the survey items is included in the **Supplementary Materials**.

Time on Schoolwork. We used the survey item probing how much of a participant's awake time they spent on schoolwork, answered from 1 = "Very little of my time (0-20%)" to 5 = "Almost all of my time (81-100%)".

Time Interacting. We used the survey item probing how much of a participant's awake time they spent interacting with others, answered from 1 = "Very little of my time (0-20%)" to 5 = "Almost all of my time (81-100%)".

Want Others. We used the survey item probing how much a participant wanted to be around others, answered from -5 = "Wanted to be completely alone" through 0 = "Neutral" to 5 = "Wanted to be completely with others".

Negative Affect. We aggregated individual survey items probing how sad, upset, hostile, irritable, and anxious participants felt that day, each answered from 1 = "Very little or not at all" to 5 = "Extremely", into a daily mean of negative affect.

Survey responses submitted between 5PM (local time) the day the survey opened and 6AM the following day were considered on time. Survey responses submitted past 6AM the day after the survey was prompted were discarded and deemed missing. A participant was included in analysis if they were compliant with at least 100 daily surveys across the data collection period, and only on-time surveys from those participants were included.

Daily Phone-Based Voice Diaries

Immediately after submitting the daily phone surveys each night, participants completed a daily voice diary via the Beiwe application. They were instructed to describe how they felt over the past 24 hours in relation to events that occurred and/or upcoming ones on their mind. As described in detail below, we used these diaries to identify specific instances of stressful events encountered by participants in daily life. Participants were asked to speak for at least one minute, and the recording automatically stopped after four minutes. To maintain privacy, participants were asked not to mention specific names of people or places, and they had the opportunity to review and re-record the voice diary before submitting. See **Supplemental Materials** for the written instructions provided to participants.

As with the daily phone-based surveys, voice diary entries submitted between 5PM (local time) the day the diary became available and 6AM the following day were considered on time. Entries submitted past 6AM the day after the diary went live were discarded and deemed missing. Additionally, audio files flagged as having unusually low volume were manually reviewed, and any empty or inaudible files were deemed unusable. Diaries that passed these two filters (time of submission and audible content) were selected for transcription (N=8,020, or 76% of total possible diaries).

Transcription of the daily diaries was outsourced to TranscribeMe, a HIPAA compliant transcription service. Voice recordings selected for transcription were sent to TranscribeMe via

secure file transfer protocols (SFTP). Transcribers were asked to transcribe the voice diaries verbatim (including filler words like "um") and to remove all personally identifying information and replace it with a "[redacted]" tag. Transcripts were shared back with us via SFTP. We spot-checked approximately 10% of the diary transcriptions for each participant against the original voice recordings. We only noticed a few instances where the transcriber had made a mistake, and in all instances, it was only one or two words that were misidentified; usually these were slang associated with college students (e.g., "PSAT" instead of "p-set", which refers to problem set, a common type of assignment in STEM classes) or terms and acronyms specific to this university (e.g., "session" instead of "section", or "path" instead of "PAF", or peer advising fellow). Any identified mistakes in the reviewed transcriptions were corrected during this spot-checking process, but we did not review nor correct any potential transcription errors in the rest of the dataset, as the small rate of errors was unlikely to cause major distortions in our downstream natural language processing analyses.

Objective Sleep Duration and Waketime Activity via Actigraphy

Sleep duration and waketime activity were derived from the accelerometer data collected through the actigraphy wristband (GENEActiv Original, Activinsights Ltd., Kimbolton, UK) and analyzed via the Deep Phenotyping of Sleep (DPSleep) processing pipeline (Rahimi-Eichi et al., 2021). As originally described in (Vidal Bustamante et al., 2022), participants wore the wristband continuously on their nondominant wrist, including during sleep and when bathing. Tri-axial acceleration was collected with a sampling frequency of 30 Hz during the academic semesters and 10 Hz during the winter break (in order to extend battery life and memory while participants were away from campus). We instructed participants to press the wristband's button when they began trying to fall sleep at night and immediately after they woke up in the morning. When on campus, participants exchanged their wristband for a fully charged one with reset memory every few weeks as part of the scheduled check-ins with the study staff.

We applied the DPSleep processing pipeline to the raw actigraphy data to detect the major sleep episode of each day. Before any analysis, the pipeline used the minute-by-minute standard deviation of the tri-axial accelerometer data to identify and subsequently remove the minutes when the individual was not wearing the device. Then, the pipeline processed the remaining accelerometer data to produce minute-by-minute activity estimates. These estimates

were computed by calculating the frequency spectrum of the accelerometer data across three axes. The intensity of the signal was represented by the area under the frequency spectrum curve for each minute. Thus, the activity estimate for each minute was quantified by the root mean square of the intensities from the three axes. Then, the pipeline estimated the major sleep episode based on a sliding window. Days where one of the boundaries of the sleep episode (i.e., rises in relative activity both before and after a period of lower activity) could not be detected due to missing data were labeled as unusable. Two trained independent raters examined the automatically detected start and end times and usability label of each sleep episode against the minute-based activity levels and the participant button presses when available. When necessary, they adjusted the automatic times and labels. A full description of the processing pipelines applied to the actigraphy data, including quality control steps, can be found in (Rahimi-Eichi et al., 2021).

All data that passed quality control was included in analysis, including days with no detected sleep episode (i.e., with no extended periods of lower relative activity). The following metrics were derived from the processed actigraphy data and used in the present analyses:

Sleep Duration. Daily sleep duration reflects the number of minutes between the estimated start and end of the day's longest detected sleep episode. Sleep duration is assigned to the day when the sleep episode ended (i.e., the day when the participant woke up).

Waketime Activity. Daily waketime activity was obtained by averaging minute-based activity, calculated as explained above, across the full duration of the wake period (i.e., the period between two detected sleep episodes in consecutive days). If daily sleep episode data was missing, the waketime duration for that day and associated physical activity were deemed missing. Given that our operationalization of waketime activity using the minute-level area under the frequency spectrum curve can yield small numbers between zero and one, we multiplied all observations by 100 to make all downstream processing and analysis more tractable.

Mental Health ("Global Clinical Symptoms") and Other Web-based Questionnaires

As originally described in (Vidal Bustamante et al., 2022), participants completed a battery of questionnaires at baseline, at the midpoint of the study period, and at the end of the study period. A physical and mental health questionnaire asked participants to report whether they had past or current diagnoses of a series of conditions listed on a checklist (**Supplementary**

Materials). We report these psychiatric diagnoses in our description of the study sample above. Additionally, participants completed the Symptoms Checklist 90 Revised (SCL-90-R; Derogatis, 1975; Derogatis & Unger, 2010), which we used for analyses. Participants were asked to complete these assessments using the REDCap online platform (Harris et al., 2009) from a computer with Internet access, in a private, quiet location and in one sitting.

The SCL-90-R is a 90-item self-report questionnaire that assesses the severity of a broad range of psychological problems and clinical symptoms, including somatization, internalizing, psychoticism, and other domains. Each question asked participants to indicate how much they were bothered by that problem during the past two weeks using a 5-point Likert scale, from "Not at all" (0) to "Extremely" (4). The Global Severity Index (GSI) is a subscale of the SCL-90-R that reflects in a single value the overall current level of distress in terms of both number of symptoms endorsed and intensity of distress (Derogatis, 1975). To provide a clearer label of what this measure represents, throughout the manuscript we refer to the GSI as "global clinical symptoms".

The GSI was computed for each participant at each of the three timepoints, transformed to adolescent- and gender-normed t-scores, and then averaged to obtain a single, person-level summary score of psychological distress across symptom domains (possible range following t-score transformation = 25-81). Symptoms are considered to be at clinically meaningful levels if the GSI t-score is > 63 (Derogatis, 1975). The GSI subscale has been found to have good sensitivity and reliability in psychiatric and non-psychiatric populations, and supported as a useful global measure of psychological distress (Holi et al., 1998; Rytilä-Manninen et al., 2016; Schmitz et al., 2000).

Analytical Approach

Manual Rating of Daily Life Stressors using Voice Diary Transcripts

The transcripts of participants' daily voice diaries (N=8,020) were used to identify the occurrence of academic and social stressful events. The rating system was developed by two of the authors (CVB and RLB) and tested over a pilot sample of four first-year college students that each provided one semester of data (444 total diaries). The master rater (CVB) rated each of the

pilot diaries. A team of four additional research assistants were trained on the rating system using this pilot dataset prior to applying the system to the target, fully independent year-long dataset.

The present rating system was modeled after the UCLA Life Stress Interview (LSI; Hammen, 1988) and the Daily Inventory of Stressful Events (DISE; Almeida et al., 2002; Almeida & Horn, 2004). Trained raters reviewed each transcript and indicated whether there was mention of a stressful event (binary 1/0) in the following life domain categories (adapted from those in the Life Stress Interview): academic (e.g., poor exam performance, heightened workload, struggling with an assignment due the next day), social (e.g., tensions or arguments with roommates, friends, romantic partner, family, or other; being rejected from joining a social club), and other (e.g., events related to health or finances). The "other" category was infrequent (<2% of all diaries and <1% between-subjects median) and excluded from the present analyses. Per the LSI and the DISE's separation of stressful *events* from *perceived* stress, raters only marked specific stressful *events* that occurred on the day the diary was recorded; they did not mark instances where participants expressed an affective response in the absence of a concrete event, occurred that same day, that triggered those feelings. For example, mentioning receiving a bad grade that day was marked as a stressful event, but we did not mark mentions of general concerns about academic performance in the absence of a concrete event occurred that same day.

In addition to the life domain of the stressful event, raters noted the severity of the event. Since we were interested in identifying events substantial enough to be potentially associated with significant changes in affect and behavior, raters were instructed to only mark events that were relatively moderate or severe, and to leave out mild hassles. The severity of a stressful event was assessed based on the relative impact of the event on an average college student's life (similar to the LSI), and inspired by DISE's definition of severity as "the degree and duration of disruption and/or unpleasantness created for the respondent" (Almeida et al., 2002; Almeida & Horn, 2004). Struggling with an assignment due the next day or receiving a bad grade on an assignment were considered moderate and severe examples of academic stressful events, respectively, as these have concrete consequences for a student's academic standing. Within social stressful events, interpersonal tensions were considered moderate, while arguments and romantic break-ups were considered severe. In contrast, merely having to do daily homework or study, or being inconvenienced by someone in passing were considered too mild to be included as stressful events. Across all participants, 4% of diaries contained a severe stressful event and

19% a moderate stressful event. We collapsed across these two categories (within academic and social domains) in our downstream data analyses for simplicity. The **Supplementary Materials** contain a description of the procedure used to train the raters and a written summary of instructions provided to them, including examples of academic and social stressful events.

All raters showed 90% or higher accuracy in the rating of the pilot diaries before proceeding to scoring the target dataset. The team of four raters split the rating of the transcripts for the target sample roughly equally, and all transcripts from one participant were assigned in full to the same rater. To ensure inter-rater reliability and consistency in the scoring within and between participants' data, the master rater scored 36 diaries for every participant: the first 12 diaries, 12 roughly in the middle of the dataset ("middle 12"), and 12 toward the end of the year ("final 12"). The master rater set up raters with each participant's dataset by going over the first 12 diaries together. The rater then scored the first half of that participant's data independently in one sitting, followed by a review of the "middle 12" diaries with the master coder; any deviations from the master rater's answers were discussed. The same procedure was repeated with the "final 12" diaries after the raters completed scoring the second half of that participant's data.

Scoring a target participant's full dataset of diaries took on average three hours, including time spent debriefing with the master rater. All daily diary data were always kept in secure, password-protected, laboratory-owned computers throughout the scoring process, including during the training phase.

Natural Language Processing of Diary Transcripts: Sentiment Analysis and Topic Modeling

Transcript Preprocessing. The only preprocessing done to the daily diary transcripts prior to fitting the sentiment analysis and topic models described below was to remove sentence timestamps and the "[inaudible]" tags and question marks added by the transcribers next to words they were unsure about (e.g., "I really enjoyed watching the [game?] with my roommates" was modified to "I really enjoyed watching the game with my roommates"). This preprocessing was done in R 4.3.1 (R Core Team, 2023).

Sentiment Analysis. Sentiment analysis was used to determine the diaries' overall valence or sentiment (negative, positive, or neutral). We used a deep learning language model ('twitter-roberta-base-sentiment-latest') freely available on Hugging Face's website

(https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest). This model uses a RoBERTa base architecture trained on over 120 million tweets (Loureiro et al., 2022) that was further finetuned for sentiment analysis (Rosenthal et al., 2019). We applied this model to our diary data to obtain automated scores (ranging from 0 to 1) for negative, positive, and neutral sentiment at the sentence level. We fit the model in Python 3.8.18 and subsequent processing was done in R 4.3.1 (R Core Team, 2023). We used the maximum of these three scores to categorize sentences as (mostly) negative, positive, or neutral in sentiment. Subsequently, we computed two transcript-level metrics: 1) we obtained the transcript-level percentages of sentences assigned to each of the three sentiment categories (e.g., a transcript could have 50% of sentences labeled as neutral, 25% as negative and 25% as positive); and 2) transcripts were categorized as (mostly) negative, positive, or neutral based on the sentiment label of the majority of the sentences in that transcript (e.g., using the previous example, that transcript would be categorized as neutral).

Topic Modeling. We identified frequent topics in our corpus of daily diaries via BERTopic v.0.15.0 (Grootendorst, 2022) using Python 3.8.18. BERTopic is a multi-layer algorithm that applies large language models and clustering techniques to create "topic representations" from text. Text was first transformed into numerical representations by using a word embeddings technique, the *all-MiniLM-L6-v2* model from Hugging Face's *sentence-transformers* package v.2.2.2 (Reimers & Gurevych, 2019). This technique projects text onto a 284-dimensional space based on the text's meaning; sentences that are close to each other in meaning are close to each other in this multidimensional space. The algorithm then reduces the dimensionality of these embeddings using a UMAP technique (Uniform Manifold Approximation and Projection) with the *umap-learn* package v0.5.4 (McInnes et al., 2020), which preserves both local and global structure of the text dataset and helps with the subsequent clustering of this space into semantically coherent topics (McInnes et al., 2020). For the clustering step, we selected an HDBSCAN clustering algorithm (Hierarchical Density-Based Spatial Clustering of Applications with Noise) due to its flexibility: it can find clusters (topics) based on the data distribution without making assumptions about the shape of the clusters and without requiring a predefined number of clusters, as well as allowing the assignment of text that does not clearly belong to any of the identified cluster to be assigned to an outlier or "junk" cluster (Campello et al., 2013; McInnes et al., 2017). The clustering was done with the *hdbscan*

package v.0.8.33 (McInnes et al., 2017). Finally, BERTopic creates "topic representations" (i.e., interpretable representations of a cluster's meaning) by identifying the terms that are most representative of each topic using a cluster-level term frequency inverse document frequency (tf-idf) matrix; words with high tf-idf scores are frequent in a cluster and infrequent across other clusters.

We applied BERTopic at the sentence-level, i.e., each sentence in each diary was assigned to one of the identified topics. In addition to the settings outlined above, we set up the model with a minimum cluster size of 100, i.e., the model was to identify topics with at least 100 representative sentences across our corpus of daily diaries. All other parameters in the HDBSCAN clustering algorithm were left in their default setting. Moreover, to leverage our prior knowledge of the content of the diaries and help with the identification of meaningful topics, we fit a semi-supervised version of BERTopic, which takes a suggested (not forced) list of topics to look for. We included a list of ten seed latent topics encompassing various domains of student life: schoolwork, people, self, health, finances, meals, sleep, exercise, housing, and extracurriculars. We exemplified each of the latent topics with four terms. For example, the following terms were used to represent the topic "schoolwork": "homework", "paper", "problem set", and "exam". For a "people" topic, the following terms were used: "friend", "girl", "guy", "parents".

Our BERTopic model identified 101 clusters or candidate topics (containing 91,251 sentences), in addition to the outlier or "junk cluster" containing all the sentences not assigned to any of the identified topics (34% of the sentences). We reviewed each cluster's representative words and sample sentences assigned to it to discern topic meaning and assign a topic label accordingly. Eight of the topics (mostly filler sentences, as explained below) were found to be redundant in meaning with other topics and were merged accordingly, resulting in a final set of 93 topics. A list of all 93 topics, their size, representative words, and sample sentences are presented in **Supplementary Table 2**.

The ten topics we suggested to the model were identified by BERTopic, although at different levels of granularity. For example, for "schoolwork", separate topics were identified for each of the four terms used as examples, whereas only one topic was identified for "people", which gathered sentences mentioning the four example terms we had suggested. About 14% of the identified topics related to students' daily schedule, 14% to subjective feelings and

assessments of their day, 10% academic and extracurricular activities, 10% social life, 8% health status, exercise, or sleep, and 3% other topics related to entertainment, grooming, chores, and financials. Additionally, 7% of the identified topics corresponded to filler sentences mentioned regularly by participants as they started and concluded their recordings and as they transitioned through different topics (e.g., "Hello voice recording", "What else happened today?", "OK, that's it.").

Bayesian Hierarchical Linear Models Assessing Same-Day and Next-Day Associations Between Stressful Life Events and Behavioral and Affective Metrics

Bayesian hierarchical linear models assessed whether the daily behavioral and affective metrics of interest (sleep duration, waketime activity, time spent on schoolwork, time spent interacting, wanting to be around others, negative affect) on day *t* varied as a function of whether an academic and/or a social stressful event was reported the same day. Separate models were fit for each dependent variable (DV) of interest. Academic and social stressful event occurrences were added as two separate binary predictors (1/0). Given the large differences in demands experienced by students over winter break (a five-week period between semesters when school is not in session and most students are off campus) we decided a priori to exclude data collected over this break from all analyses. Moreover, given weekly fluctuations in the DVs within the semesters, day of the week was included as a covariate in our models. Random intercept and slopes for stressful events were added by participant, and the model assumed a normally distributed random error term. The regression model was specified as follows:

$$DV \sim Academic\ Stressful\ Event + Social\ Stressful\ Event + Day\ of\ Week + (1+Academic\ Stressful\ Event + Social\ Stressful\ Event \mid participant)$$

While same-day effects were the primary focus, we also assessed whether the association between stressful events and the affective and behavioral metrics of interest would extend to the next day. To assess the association between today's affective and behavioral metrics and yesterday's stressful events, we used the same model as above, but lagged the stressful events by one day (i.e., on day *t-1*).

Note that sleep duration was the only tested DV that occurred at naturally lagged intervals relative to the stressful events (i.e., events only occurred during waketime). In the "same-day" analysis, we report the association between a stressful event and the duration of the

sleep episode that ended the morning of the event (i.e., the sleep episode that preceded the stressful event), while in the "next-day" analysis, we report the association between a stressful event and the duration of the sleep episode that started later that night (i.e., the sleep episode that followed the stressful event).

Bayesian models were estimated in R 4.3.1 (R Core Team, 2023) using the Stan modeling language (Stan Development Team, 2018) and the packages *rstanarm* v2.21.4 (Goodrich et al., 2023), *tidybayes* v3.0.1 (Kay, 2021), and *bayestestR* v0.13.1 (Makowski, Ben-Shachar, & Lüdecke, 2019). All models were fit with a gaussian distribution. Posterior predictive checks indicated that the models with Likert-scale DVs (i.e., the survey items) may be better specified as ordinal regressions with a logit link function, but we opted for gaussian distributions across all models to facilitate comparison of effect estimates. Bayesian models were fit using MCMC with the default weakly informative priors from rstanarm (i.e., rescaled N(0, 2.5) priors for the slope parameters). For each parameter, we sampled from four stationary Markov chains, each comprising 5,000 sampling iterations, including a burn-in period of 2,500 iterations that were discarded, for a total of 10,000 post-warmup draws. Convergence of the four chains to a single stationary distribution was assessed quantitatively via the R-hat convergence diagnostic (Gelman & Rubin, 1992), with adequate convergence defined as R-hat < 1.1, and qualitatively by visual inspection of trace plots showing the estimated parameter as a function of each chain's iteration number, with adequate convergence defined as the chains overlapping with each other throughout and a lack of structured patterns in each chain. All models showed adequate convergence. We also report each model's Effective Sample Size (ESS), which generally should be at least 1,000 to obtain stable estimates (Gelman et al., 2013; Muth et al., 2018), as was the case with all our models.

Point estimates of intercepts and slopes were computed as the median value of their respective posterior distributions. Ninety-five percent uncertainty intervals (UI) were computed as the 2.5% and 97.5% quantiles of the posterior distribution. To provide intuitive parallels to a frequentist inference framework, we interpreted a predictor slope as "statistically significant" if its 95% UI did not contain zero, or put differently, if the proportion of the posterior distribution falling in the direction (positive or negative) of the point estimate (also known as "probability of direction", or pd) was higher than 0.975, which approximates a frequentist one-tailed p-value < 0.025 (Makowski, Ben-Shachar, Chen, et al., 2019).

Exploratory Associations between Global Clinical Symptoms and Stressful Event Frequency, Topic Frequency

An exploratory between-participants linear regression assessed whether participants' year-long frequency of academic and social stressful events was associated with their mean global clinical symptom scores (see description earlier in the Method section). The outcome variable was participants' global clinical symptoms averaged across the three timepoints, and the two predictor variables were the year-long frequency of academic stressful events and the frequency of social stressful events (both computed out of the total number of diaries available for each participant). This model was estimated in a Bayesian framework using the same approach described in the previous section for the hierarchical models.

Finally, two exploratory correlations assessed the relationship between the participants' mean global clinical symptom scores and the person-level proportion of mentions of the topics "schoolwork" and "people" classified as negative in sentiment per the results of the natural language processing described earlier. These correlations were estimated in a Bayesian framework using the package *BayesFactor* v0.9.12.4.7 (Morey & Rouder, 2024).

Results

Every night for a full academic year, participants submitted voice diaries narrating the main events of their day. Additionally, they submitted daily surveys and wore an actigraphy wristband throughout the study period. **Table 1** presents between-person summary statistics for the main variables used in analysis.

Participants provided a median of 184 usable voice diaries (range 70-221 out of a total possible of 223), and diaries were on average 1.60 minutes (256 words, 17 sentences) long. Participants' total number of usable diaries was not correlated with their mean negative affect reported in daily surveys (r [median of posterior distribution] = .10, 95% UI = [–.36, .17], pd = .78), and participants' daily negative affect did not differ as a function of whether they submitted a voice diary that day or not (b [median of posterior distribution] = –0.02, 95% UI = [–0.10, 0.06], pd = .70). These observations suggest that participants' daily affective state was not systematically associated with whether they submitted a usable voice diary (more detail on available data by participant and data stream is included in **Supplementary Table 1**).

Table 1. Between-person summary statistics of daily metrics.

	Usable Total	Usable per Part. (Med.)	Mean	Median	Min.	Max.	SE
Diaries: Length	8,280	184	-	-	-	-	-
Voice file (Minutes)	-	-	1.60	1.39	1.02	3.19	0.08
Transcript (Word Count)	-	-	256.01	227.36	138.32	510.91	13.42
Transcript (Sentence Count)	-	-	16.53	15.81	8.93	28.90	0.52
Diaries: Stressful Events	8,280	184	-	-	-	-	-
Neither (%)	-	-	77.51	78.67	56.80	93.61	1.32
Academic (%)	-	-	13.24	12.04	3.57	33.33	0.87
Social (%)	-	-	10.17	8.33	0.46	29.36	1.03
Both (%)	-	-	0.91	0.63	0	2.80	0.13
Diaries: Sentiment	8,280	184	-	-	-	-	-
Negative (%)	-	-	25.40	25.79	9.56	46.24	0.98
Neutral (%)	-	-	43.74	41.95	28.95	66.59	1.34
Positive (%)	-	-	35.87	37.75	20.19	52.85	1.16
Sleep Duration (Minutes)	9,490	199	431.12	433.63	359	497.65	4.79
Waketime Activity (Units)	9,466	199	4.98	4.86	3.12	8.04	0.19
Time on Schoolwork (1 to 5 Likert)	8,807	193	2.58	2.53	1.39	4.06	0.06
Time Interacting (1 to 5 Likert)	8,807	193	3.31	3.31	1.65	4.62	0.07
Want Others (-5 to 5 Likert)	8,807	193	1.49	1.28	-1.94	4.09	0.18
Negative Affect (Mean of 5 Items, 1 to 5 Likert)	8,807	193	1.76	1.66	1.11	2.77	0.06

Part. = Participant; Med. = Median; Min. = Minimum; Max. = Maximum; SE = Standard error of the mean.

Content of Students' Daily Voice Diaries Fluctuated Along with the Academic Calendar

The content of the students' daily voice diaries fluctuated as expected with the main landmarks in the first-year student's academic calendar (**Figure 1**), serving as a validation of

participants' sustained compliance and our ability to capture real-world dynamics relevant to students' academic and social life.

A deep learning natural language processing algorithm identified common topics mentioned in the diaries across participants, four of which are highlighted in **Figure 1** as a means to establish face validity. Mentions of "midterm exam" were most frequent in the weeks half-way through the Fall and Spring Semesters, when most university courses schedule midterm assignments and examinations. Meanwhile, mentions of "final exam" were most frequent during final exams period, the two weeks at the end of each semester when courses hold their final examinations, as well as during reading period, the week leading up to the final exams when no classes are held and students prepare for their exams. The fact that recall of these topics is observed not only in the fall but also in the spring semester demonstrates participants' continued compliance through the end of a full year of daily study participation. Moreover, mentions of both topics were nearly absent at other times in the year, demonstrating specificity in the mentions.

Another example of good recall and specificity in topic mentions comes from a special feature of first-year students' social life. On housing day, an annual university event held a couple of days before spring break, first-year students and their friends are assigned to one of the university's residential houses, where they will live for the remainder of their college years. Weeks ahead of housing day, students must form a "blocking group" with the friends they wish to enter the housing lottery with. These groups are capped at eight, meaning that students must make collective decisions about who to include and exclude from their group. In line with these events, increased mentions of "blocking group" led up to housing day and mentions of "housing" spiked on housing day, while remaining close to zero the rest of the year.

Together, these patterns reinforce that the daily diaries captured real-world events and suggest that we are well positioned to study naturalistic instances of both academic and social stress in students' daily life.

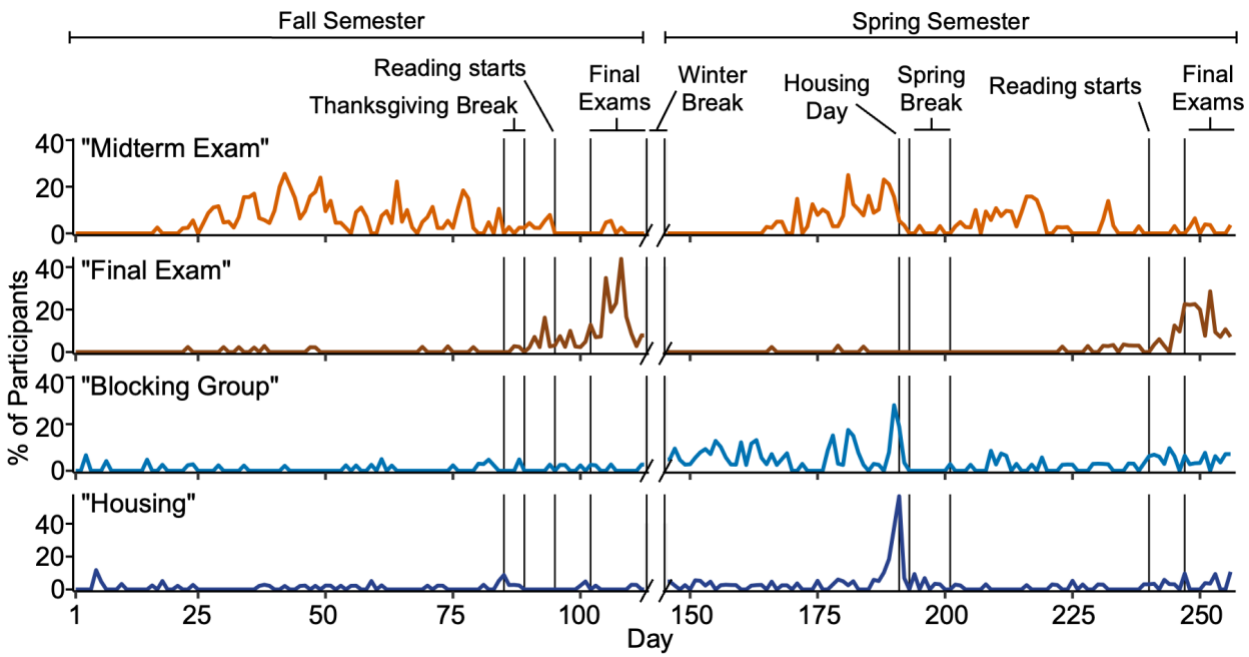

Figure 1. Students' daily voice diary contents showed expected patterns across the academic year. Time series show the percentage of participants that mentioned each of the topics in quotations each day in the academic year, as identified through a deep learning topic modeling algorithm applied to the full corpus of students' voice diary transcripts. Note topics increase their frequency selectively at appropriate times during the year. Black vertical lines indicate important dates in the school calendar, labeled at the top.

Daily Diaries' Diverse Sentiment and Topics Revealed the Richness of Students' Daily Life

The daily voice diaries instructed participants to speak between one to four minutes about the main events of their day and how they felt about them but were otherwise unstructured. The wide variety of sentiment and topics expressed in the diaries demonstrates that participants felt comfortable sharing diverse aspects of their daily life, including both the good and the bad (see **Table 2** for example sentences with neutral, positive, and negative sentiment for the topics highlighted in **Figure 2**).

Students shared a mix of neutral, positive, and negative sentiment expressions in their daily diaries. **Figure 2a** shows results of deep learning natural language processing algorithms applied to analyze the sentiment of the students' daily diaries. On average, most sentences within a diary were neutral (44%) or positive (36%) in sentiment (**Figure 2a**, left), e.g., descriptions of students' daily routines and expressions of classes and/or social interactions they enjoyed. Although negative sentences (e.g., struggling with class material or social life, expressions of being sick or in a negative mood) were more rare, they represented a non-trivial amount of each

diary's content (25%), with an average of one in every four sentences categorized as negative. When categorizing sentiment at the diary level (**Figure 2a**, center), an average of 18% of each student's diaries had a majority of negative sentences, suggesting that almost one in every five days might have been especially challenging for participants.

Figures 2b and **2c** present extended results from the topic modeling analyses described earlier (see **Supplementary Table 2** for a list of all 93 identified topics, including sample sentences and frequency for each). Overall, the range, frequency, and sentiment of the identified topics reflect the nature of college life, including both the challenges and excitement associated with both academic and social experiences.

Students' daily diaries covered a wide range of topics, spanning daily routine, (I) students' subjective feelings and assessments of their day, (II) health status, exercise, and sleep, (III) academic and extracurricular activities, (IV) interpersonal relations and social activity, and (V) entertainment and daily chores (**Figure 2b**, top). Some of the most common topics mentioned by participants were "people", "schoolwork", and "stress", reinforcing again that the daily diaries captured core aspects of the student experience, including both social and academic domains of life stress.

Topic mentions themselves varied in sentiment (**Figure 2c**), suggesting that the same topic could be mentioned in neutral, positive, or negative terms between participants as well as within participants over time. Topics like "sleep", "schoolwork", and "paper" were most often mentioned with a neutral sentiment (e.g., statements of their sleep schedule or spending time on an assignment), with roughly equal rates of negative and positive mentions (e.g., expressing dis/satisfaction with their sleep or progress made towards an assignment). As reflected by our "happy/unhappy" topic label, some of the topics related to participants' emotional states included sentences with either positive or negative references to that feeling. Other topics were more consistently associated with negative sentiment, such as "stress" and "health". Topics like "party" and "people" were most often mentioned with a positive or neutral sentiment, though the percentage of negative mentions was not trivial (41% and 19%, respectively), especially considering that "people" were mentioned in roughly half of a student's diaries.

The sentiment and topics discussed by participants showed expected regularities that reflected the weekday/weekend school schedule. Diaries tended to be less negative and more positive on Thursdays, Fridays, and Saturdays (**Figure 2a**, right). Similarly, mentions of topics

related to “schoolwork” and “stress” were least frequent on those days, while mentions of “people” and of “party” were most common on Fridays and Saturdays (**Figure 2b**, bottom).

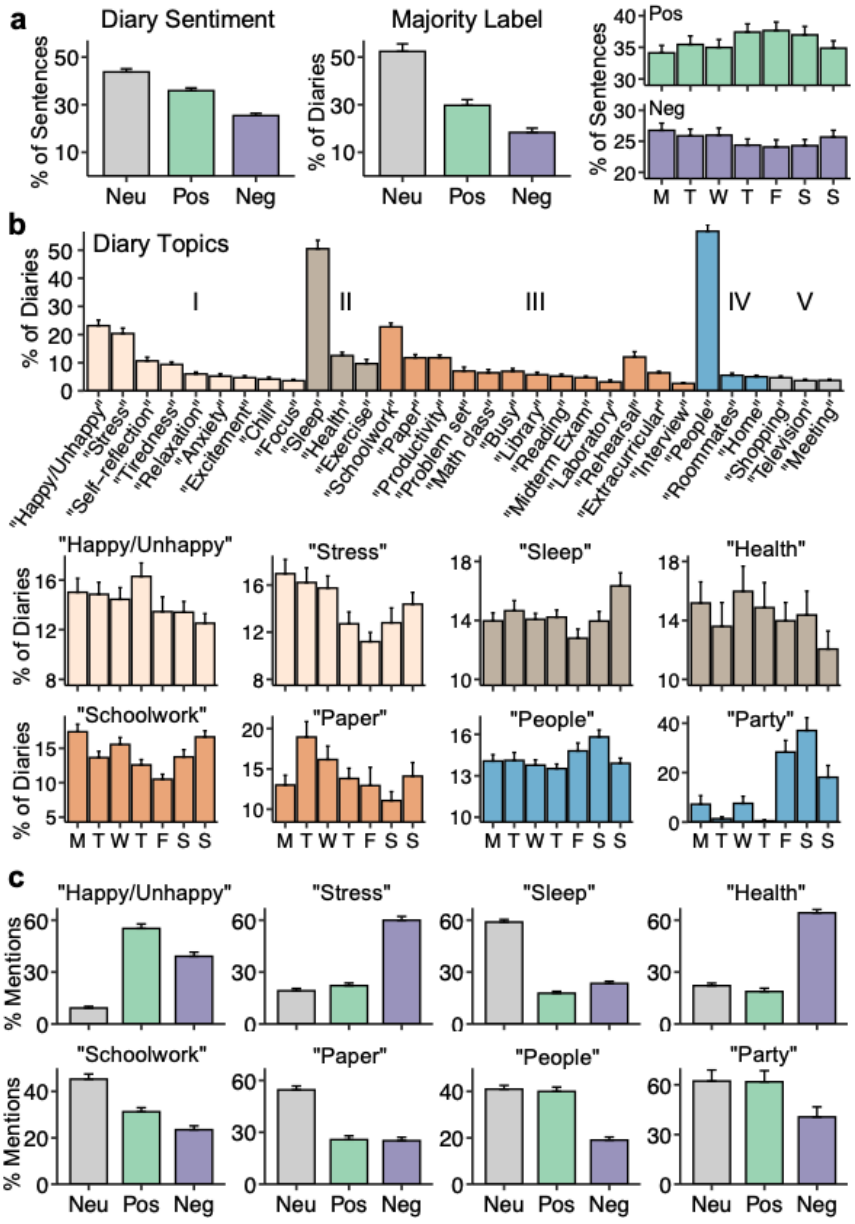

Figure 2. Variety of sentiment and topics identified in the voice diaries revealed the richness of students’ daily life. a) Diary Sentiment. *Left*: Between-participant mean percentage of sentences within a diary scored as negative, neutral, or positive in sentiment, as estimated by a deep learning model fine-tuned for text sentiment analysis. *Center:* Between-participant mean percentage of diaries that had a majority of neutral, positive, and negative sentences. *Right*: Between-participant mean percentage of sentences within a diary scored as negative (top) or positive (bottom) by day of the week (M=Monday to S=Sunday). Error bars show the standard error of the mean. **b) Diary Topics.** *Top*: Between-participant mean percentage of diaries that included mention of the topics on the x-axis, as identified by a deep learning model fine-tuned to

identify recurring topics in text. Error bars show the standard error of the mean. *Bottom*: Between-participants mean percentage of all topic mentions by day of the week for some of the most frequent topics. Error bars show the standard error of the mean. **c) Topic Sentiment.** Between-participant mean percentage of topic mentions scored as neutral, positive, and negative in sentiment. Error bars show the standard error of the mean.

Table 2. Sample diary sentences assigned to topic and sentiment categories by deep learning natural language processing algorithms.

Topic	**Sentiment**	**Sentence**
Happy/ Unhappy	Neutral	"And they made me feel better but also worse at the same time."
		"That's, like, the big message, but, like, I've been trying to think, like, what gives me joy?"
	Positive	"And let me just say that I am so happy with everything, with absolutely everything."
		"I was very excited and happy all day long, um, which of course, was very nice."
	Negative	"And, like, not knowing who we are and just, like, being unhappy with who we've become and-- 'cause at this point, I really am just not happy."
		"I'd rather just feel happy, but it's impossible to feel happy when you love someone and they toss you aside."
Stress	Neutral	"Um, so I'm preparing myself mentally for that and trying not to stress, or rather, take them one thing at a time."
		"Um, again, not really stressed about it."
	Positive	"So today was another pretty good stress-free day."
		"So today was a pretty good day overall, extremely low stress, which was nice."
	Negative	"Today was a really, really incredibly stressful day, um, just, like, tons and tons of work, just does not stop, actually."
		"But yeah, that was very upsetting and stressful, the whole, like, event of it so that was definitely a low point in my day."
Sleep	Neutral	"So today, I woke up at around 12:00 again."
		"I laid down to close my eyes for a minute, and then I woke up to my alarm the next morning."
	Positive	"I slept in this morning, so that was super cool."
		"Today started with a, uh, good night's sleep, probably my best night's sleep in a while."
	Negative	"And I really hate to lose sleep, um, on a Sunday 'cause it's just a horrible way, uh, to go into the week, being sleep-deprived and having no energy, not really wanting to take on the week."
		"I unfortunately only got two hours of sleep because I make bad life choices."

Health	Neutral	“So I might try and run by the store tomorrow if I have time to get some honey and some medicine and maybe some more cough drops and maybe some soup.”
		“I went to the doctor and got it checked out.”
	Positive	“I had a meeting with a doctor that I had set up to meet with, and turns out-- and it turned out to be the best thing ever.”
		“Today, I woke up and I was feeling quite rested, and I was actually feeling a lot better, so I think I'm definitely recovering from my cold.”
	Negative	“Um, today, my allergies have really been acting up, and so that's definitely the worst part of my day.”
		“Like, my head hurt really bad, and I just felt weird and funny, and my throat really hurt, and I was also losing my voice today.”
Schoolwork	Neutral	“Um, then after that, I continued to do work.”
		“Um, I'm gonna go to that, knock off some work.”
	Positive	“Today, I got a lot of work done, which is great.”
		“Um, I was very productive and got, like, a ton of work done, so I'm very happy about that.”
	Negative	“And then I just like didn't get anything done today, and that was like-- I just really was mad at myself because I just like didn't get anything accomplished.”
		“So now I just have this whole mess of work that I didn't do and that I really, really wish I did.”
Paper	Neutral	“Um, then, uh, I went at 5:30 to go, uh, speak to my preceptor about my paper.”
		“And now, I'm thinking about changing one of my topics for the paper that I have to write for my writing class.”
	Positive	“And I found out that I got an A on my paper, which I was really surprised about and super, super happy.”
		“Um, I got some good work done on my essay that's due Friday, so I'm really happy about that.”
	Negative	“And then I had, uh, a meeting with one of my teachers about one of my essays, and she basically told me that, like, all of it was wrong, so I have to rewrite it, which sucks.”
		“Um, and it just took me so long, and it was so frustrating because it was, like, writer's block, but also, like, the research paper has, like, unique challenges to writing it that I have found very difficult.”
People	Neutral	“I might text them and see whether they wanna hang out tomorrow or something.”
		“Um, then my little sister and my parents and my little brother were here.”
	Positive	“Today was Thanksgiving day, so spent a lot of time with family, uh, which was awesome.”

		"So I really love her a lot and actually just super, super happy to have her here and to show her the school because I really like this school and I really like my girlfriend."
	Negative	"I'm just having a hard time connecting with them, and they were totally not interested in listening, so it's kind of frustrating."
		"Seeing that person makes me extremely angry to my very core, um, and I do not wanna have any s-- any form of relationship with this kinda per-- with this person, um, so I promptly left, um, just in all fury."
Party	Neutral	"And so there were, like, no parties happening."
		"I think it was really, like, this weekend, like, I didn't do any work this weekend because my cousin was here, and we went to a bunch of parties."
	Positive	"The after-party was literally the m-- best night of my life, um, up to this point."
		"The party was absolutely fantastic."
	Negative	"Um, I went to a party and it was just not fun and, like, kind of annoying just social space to be in, and I think that, like-- I'm bringing that energy into the next day."
		"And the party was pretty cramped and kinda bad, honestly."

Academic Stressful Events Were More Frequently Reported Than Social Stressful Events

A core goal of our study was to identify and characterize the occurrence of academic and social stressful events in students' daily life. A team of trained human raters went through each of the students' daily diaries and identified mention of academic and social stressful events that occurred the same day the diary was submitted. Note that this rating system was designed to only capture mentions of specific *events*, rather than mentions of perceived stress in the absence of a concrete event that occurred that same day, and only included events of moderate to high severity (e.g., receiving a bad grade, having an argument with someone), rather than also including mild hassles (e.g., having to do daily homework, being annoyed by someone's voice in the library).

On average, academic and/or social stressful events were identified in a median of 21% of a student's diaries, or approximately one in every five days. Academic stressful events were identified in a median of 12% of a student's diaries (i.e., at a rate of about once per week), while Social stressful events were identified in a median of 8% of a student's diaries (i.e., at a rate of about once every two weeks; **Figure 3a**). A small fraction of diaries (0.6%) contained mentions

of both academic and social stressful events. Academic stressful events were more common earlier in the week (when students are more likely to attend classes) and least likely on Fridays and Saturdays (**Figure 3b**). Meanwhile, days with social stressful events did not follow the weekly school schedule as clearly.

Of note, the frequency of both academic and social stressful events varied substantially between individuals (**Figure 3a)**. For example, one participant reported social stressful events as infrequently as 0.5% of their 219 diaries, while another reported them as frequently as 29.4% of their 218 diaries, reflecting the diversity of the first-year college experience lived by different students.

Across participants, diaries were predominantly neutral in sentiment (~42% of sentences) regardless of whether a stressful event was mentioned, but diaries containing mention of either an academic or a social stressful event had a greater percentage of negative sentences (~33%) and lower percentage of positive sentences (~28%) than diaries with neither type of event (23% and 38%, respectively; **Figure 3c**). Of note, diaries containing academic or social stressful events were roughly equal in their percentage of negative sentences (~33%), suggesting that students were not consistently speaking a greater number of negative sentences when describing instances of one or the other domain of life stress.

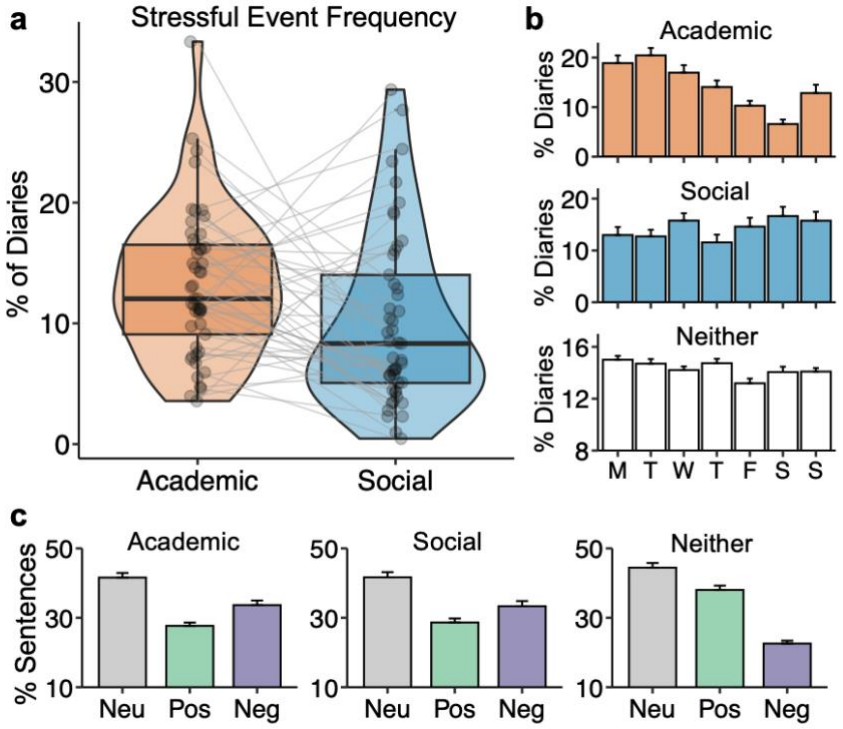

Figure 3. Academic stressful events were more frequently reported than social stressful events. a) Between-person percentage of daily diaries that contained mention of an academic or social stressful event, as identified by trained human raters. Gray circles united by light gray lines represent individual participants. **b)** Between-person mean percentage of diaries with and without a stressful event by day of the week. Error bars show standard error of the mean. **c)** Between-participant mean percentage of sentences within a diary scored as neutral, positive, or negative in sentiment, computed separately for diaries that contained mention of academic stressful events, social stressful events, or neither event type. Error bars show standard error of the mean.

Students' Behavior and Affect Showed Weekly Fluctuations Associated with the School Schedule

Ultimately, our study sought to examine how the experience of academic and social stressful events related to changes in students' daily behavior and affect captured through actigraphy wristbands and daily phone surveys. These behavioral and affective metrics showed weekly fluctuations associated with the school schedule (**Figure 4a**). On Fridays and Saturdays, participants slept longer at night, were more physically active, and spent less time on schoolwork and more interacting with others compared to the rest of the week. These patterns reinforced our decision to account for systematic day-of-week behavioral fluctuations in our statistical models of the association between stressful events and students' daily affect and behavior.

Relatedly, there were intuitive patterns of within-person correlations among the daily metrics of behavior and affect (**Figure 4b**). For example, on days when students spent more time on schoolwork, they also spent less time interacting and they were less physically active. On days when students reported wanting to spend their time around others they also spent more time interacting and reported feeling less negative affect. Additionally, self-reports of higher negative affect in the daily surveys correlated with higher negative sentiment scores in the language used by students in their daily voice diaries, suggesting that students were completing surveys and diaries conscientiously.

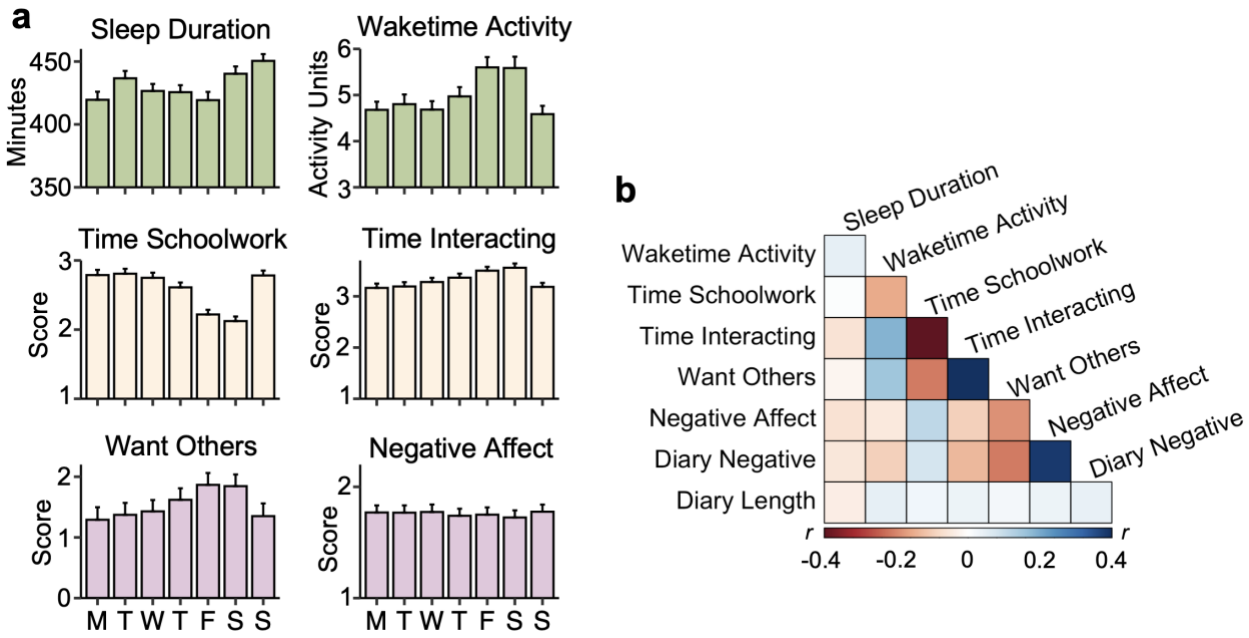

Figure 4. Behavioral and affective metrics measured through actigraphy wristbands and phone-based surveys showed weekly fluctuations. a) Between-person means of behavioral and affective metrics by day of the week. Error bars show standard error of the mean. Green indicates behavioral variables collected via actigraphy wristbands, cream indicates behavioral variables collected via daily surveys, and pink indicates affective variables collected via daily surveys. **b)** Correlation matrix shows within-person *r* correlation coefficients of daily behavioral, affective, and linguistic metrics.

Academic Stressful Events and Social Stressful Events Showed Distinct Associations to Daily Behavior and Affect

Bayesian hierarchical models estimated within-person changes in daily affective and behavioral metrics as a function of whether an academic stressful event and/or a social stressful event occurred that day (both binary predictors included within the same model), while controlling for day of the week (**Figure 5** and **Table 3**).

Compared to days with no academic stressful events, days with these events were associated with shorter sleep duration, decreased waketime physical activity, reduced wanting to be around others, and increased negative affect. Academic stressful events were also associated with reduced social interaction and increased time spent on schoolwork, above and beyond days with social stressful events (non-overlapping UIs, highlighted with asterisks in Figure 5).

Days with social stressful events stood out by their association to heightened self-reported negative affect, above and beyond that associated with academic stressful events. Other

than reduced time spent on schoolwork, we did not identify changes in sleep, physical, or social behavior as a function of whether a social stressful event occurred that day.

To explore the durability of these event-related associations, a separate set of models tested lagged associations between the occurrence of an academic and/or a social stressful event and the behavioral and affective metrics captured the following day (results shown in faded shading in **Figure 5**; **Table 3**). We found the same overall pattern of results as the same-day models except that the magnitude of the effects was attenuated.

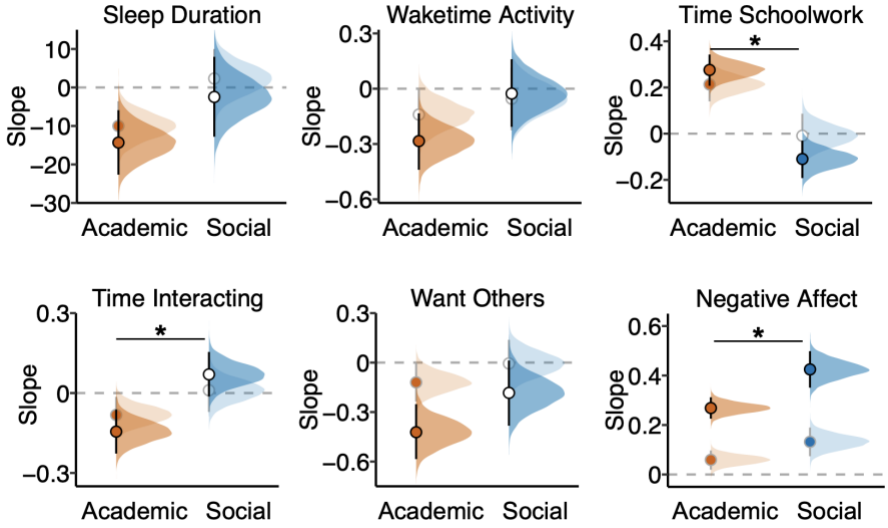

Figure 5. Academic stressful events were associated with shorter sleep duration, reduced waketime physical activity, increased schoolwork, and reduced social interaction, while social stressful events were associated with especially heightened negative affect. Results of Bayesian hierarchical linear models assessing levels of each of the title variables as a function of whether an academic (in red) and/or a social stressful event (in blue) occurred the same-day (dark shading) or the previous day (faded shading). Circles show the models' estimated slopes (unstandardized), computed as the median of the posterior distribution; colored circles denote effects that would be considered significant in a frequentist framework; white circles denote non-significant effects. Error bars show 95% uncertainty intervals (UIs). Shaded density plots show the full posterior distributions of the slopes. Asterisks (*) highlight instances where the UIs around the effects of academic and social stressful events do not overlap with each other.

Table 3. Results of Bayesian hierarchical linear models assessing behavioral and affective metrics associated with same-day stressful events (*t*) and with previous-day stressful events (*t-1*).

Model DV	Stressful Event Predictors	Slope (Med.)	Std. Slope (Med.)	95% UI	pd	R-hat	ESS
Sleep Duration	Academic (*t*)	−14.32*	−0.05*	[−22.65, −5.91]	1	1	6437
	Social (*t*)	−2.46	−0.01	[−12.79, 7.99]	.68	1	5639
	Academic (*t-1*)	−10.00*	−0.04*	[−16.42, −3.57]	1	1	12101
	Social (*t-1*)	2.33	0.01	[−5.32, 10.05]	.73	1	10603
Waketime Activity	Academic (*t*)	−0.28*	−0.04*	[−0.44, −0.13]	1	1	5154
	Social (*t*)	−0.03	−0.00	[−0.21, 0.16]	.61	1	8408
	Academic (*t-1*)	−0.14	−0.02	[−0.29, 0.00]	.97	1	10824
	Social (*t-1*)	−0.05	−0.01	[−0.23, 0.11]	.74	1	9921
Time Schoolwork	Academic (*t*)	0.28*	0.09*	[0.21, 0.34]	1	1	13137
	Social (*t*)	−0.11*	−0.03*	[−0.19, −0.03]	1	1	12760
	Academic (*t-1*)	0.21*	0.07*	[0.14, 0.29]	1	1	16285
	Social (*t-1*)	−0.01	0.00	[−0.11, 0.09]	.58	1	11540
Time Interacting	Academic (*t*)	−0.14*	−0.05*	[−0.23, −0.06]	1	1	4223
	Social (*t*)	0.07	0.02	[−0.01, 0.15]	.95	1	6644
	Academic (*t-1*)	−0.08*	−0.03*	[−0.15, −0.01]	.99	1	10859
	Social (*t-1*)	0.01	0.00	[−0.07, 0.09]	.60	1	9603
Want Others	Academic (*t*)	−0.42*	−0.07*	[−0.58, −0.25]	1	1	6476
	Social (*t*)	−0.18	−0.03	[−0.38, 0.01]	.97	1	5357
	Academic (*t-1*)	−0.12*	−0.02*	[−0.24, 0.00]	.98	1	14329
	Social (*t-1*)	−0.01	−0.00	[−0.15, 0.14]	.53	1	12117
Negative Affect	Academic (*t*)	0.27*	0.14*	[0.23, 0.31]	1	1	7616
	Social (*t*)	0.42*	0.19*	[0.35, 0.50]	1	1	4461
	Academic (*t-1*)	0.06*	0.03*	[0.02, 0.10]	1	1	10147
	Social (*t-1*)	0.13*	0.06*	[0.07, 0.19]	1	1	8306

Notes: Models were fit separately for each dependent variable (DV), and within each DV, two models were fit, one with the predictors at time *t* and one with the predictors at time *t-1*. All models included day of the week as a covariate. Med. = Median of the posterior distribution; Std. = Standardized; UI = uncertainty interval; pd = probability of direction; ESS = effective sample size. Asterisks (*) next to slope value denote effects that would be considered significant in a frequentist framework.

Frequent Occurrence of Social Stressful Events Correlated with Higher Symptoms of Psychopathology

Not all students experienced the same frequency of stressful events, but event-related rises in negative affect were shared across the sample. **Figure 6a** highlights the time series of negative affect reported by four individual participants (P1-P4) across the academic year, while the occurrence of academic and social stressful events is denoted by the background color (red background indicates an academic stressful event, and blue indicates a social stressful event). Although these four participants vary in the frequency of events, all show increases in negative affect on days with a social stressful event.

Frequent occurrence of social stressful events was associated with higher symptoms of psychopathology, reported at baseline and at the end of each semester through a clinical questionnaire assessing global clinical symptoms tied to psychological distress. P1 and P2, who had infrequent social stressful events, scored low on global clinical symptoms, while P3 and P4, who had frequent social stressful events, scored high on global clinical symptoms (**Figure 6b**). In line with these observations, an exploratory between-participants linear regression model including the full sample of participants found that the frequency of social stressful events (*b* [median of posterior distribution] = 49.69, 95% UI = [13.97, 84.27], pd = 1), but not the frequency of academic stressful events (*b* [median of posterior distribution] = 25.98, 95% UI = [–14.74, 67.27], pd = .90), was associated with higher global clinical symptoms.

Additional exploratory correlations leveraging the sentiment of the language used by students in their daily diaries revealed similar patterns of results (**Figure 6d**). Students that had a higher proportion of mentions of other "people" using negative sentiment language had higher global clinical symptoms (*r* [median of posterior distribution] = .51, 95% UI = [.25, .68], pd = 1). We did not find a correlation between the proportion of negative mentions of "schoolwork" and global clinical symptoms (*r* [median of posterior distribution] = .15, 95% UI = [–.12, .40], pd

= .87). Together, these exploratory findings suggest that frequent negative interpersonal experiences in daily life may be a marker of enduring psychological distress.

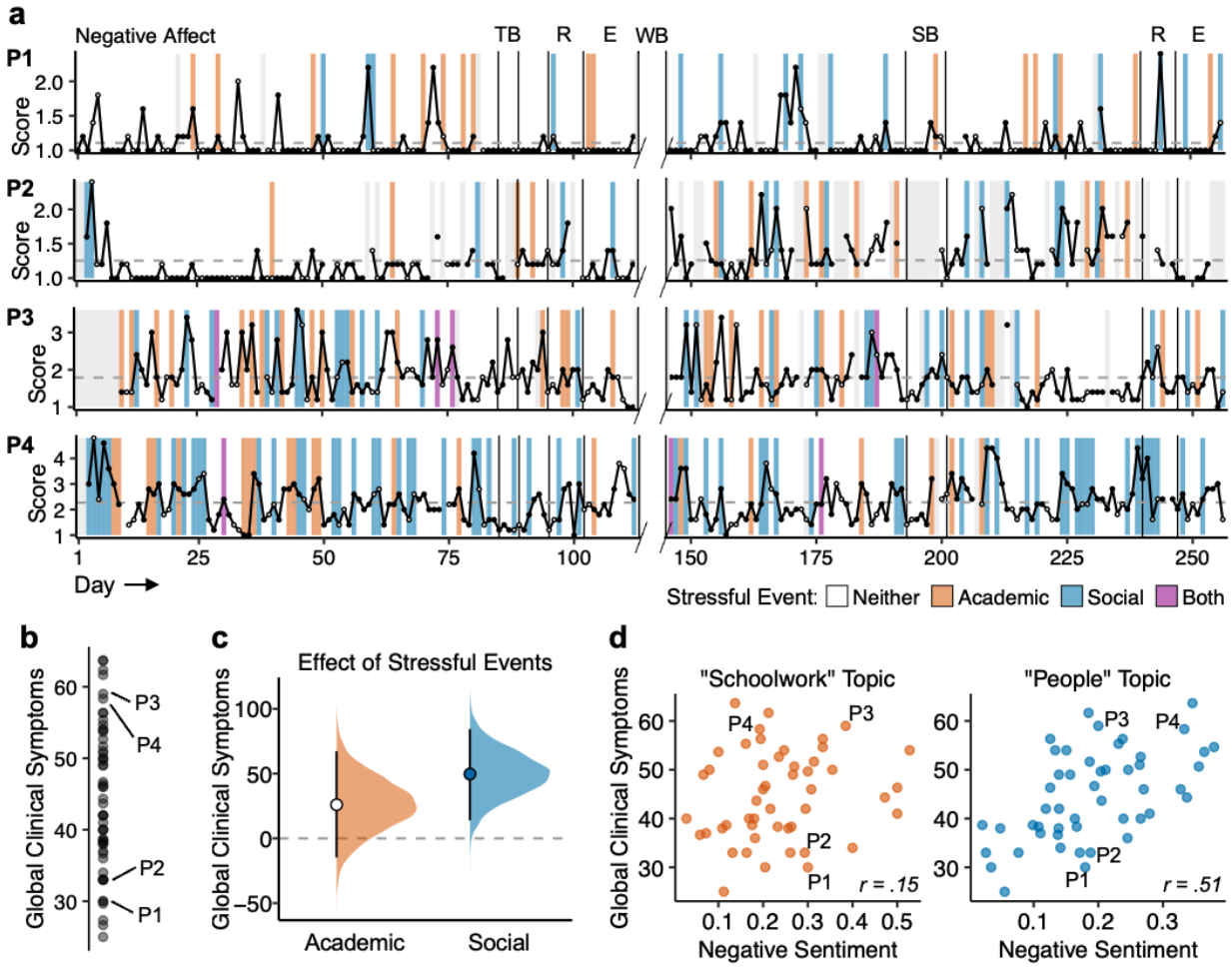

Figure 6. Social stressful events were associated with increased same-day negative affect across participants, and their frequent occurrence was correlated with higher symptoms of psychopathology. a) Time series of daily self-reported negative affect for four representative participants (P1 to P4) over the course of the academic year. Background is shaded according to whether a stressful event occurred that day (white signals neither academic nor social event occurred; red signals an academic stressful event, blue signals a social stressful event, and dark magenta signals that both an academic and a social event occurred that day). Light gray shading shows missing negative affect data. Dashed gray horizontal line shows the participant's mean negative affect. Black vertical lines indicate important dates in the school calendar. TB = Thanksgiving break, R = reading period, E = exams period, WB = winter break (excluded from analyses), SB = spring break. **b)** Between-participants distribution of global clinical symptoms. Each dot represents a participant. The four participants highlighted in panel (a) are labeled for reference. **c)** Results of a between-participants linear regression model assessing participants' mean global clinical symptoms as a function of the frequency of academic stressful events and the frequency of social stressful events occurred over the course the school year. Circles show the models' estimated slopes, computed as the median of the posterior distribution; the colored

circle denotes an effect that would be considered significant in a frequentist framework, and the white circle denotes a non-significant effect. Error bars show 95% uncertainty intervals. Shaded density plots show the full posterior distributions of the slopes. **d)** Between-participants associations between the mean global clinical symptoms and the proportion of mentions of "schoolwork" (left) and of "people" (right) topics that used language classified as negative in sentiment per natural language processing algorithms. *r* coefficients represent the median of the respective posterior distribution.

Discussion

Stress has been consistently linked to poor mental health outcomes, but less is known about the more proximate (day to day) impacts of daily life stress. The current study collected actigraphy data, daily surveys, and daily voice diaries continuously over an academic year to examine first-year students' experiences of academic and social stressors and their association to affect and behavior. Group-level models assessing within-person associations found that days with academic stressful events were characterized by shorter sleep duration, reduced physical activity and social interaction, increased time spent on schoolwork and modest increases in negative affect. Meanwhile, days with social stressful events stood out by their association to especially heightened negative affect, above and beyond the effect of academic stressful events. Together, these results echo and extend prior research on daily life stress in multiple ways.

While researchers began studying individuals' daily life stress decades ago via pen-and-paper diaries (Bolger et al., 2003; Tennen et al., 1991), our work demonstrates that recent technological and analytical advances have enabled the characterization of daily life with unprecedented granularity and over extended periods of continuous examination (Vidal Bustamante et al., 2022). Cutting-edge natural language processing techniques helped unveil and quantify the wide range of topics and sentiment shared by students in their daily voice diaries, suggesting that participants were comfortable sharing both positive and negative aspects of their daily life, and reinforcing the suitability of our intensive longitudinal data for the investigation of academic and social stressors in the real world.

Most critically, we build on prior literature by directly assessing multiple behavioral and affective markers of different domains of life stress within individuals. The different patterns of effects identified for academic and social stressful events may be related to academic stress being more common and structured than social stress, as reflected by its higher frequency rates and systematic weekly patterns observed in Figure 3. When workload increases and an important assignment is due, a common response is to devote time to that task, sometimes at the expense of

social interaction, physical activity, and even sleep. In contrast, social stressful events like interpersonal conflicts are not task-like; they have less of a concrete, intuitive strategy to deal with the situation in the immediate term. In this case, behavioral responses are likely to be more idiosyncratic to the situation and context, as well as to the individual and their preferred coping strategies (DeLongis & Holtzman, 2005; Folkman & Moskowitz, 2004; Lee-Baggley et al., 2005). This heterogeneity would make it more unlikely to find stable, group-level behavioral effects associated to social stressful events.

Nevertheless, days with social stressful events stood out by their heightened negative affect (the largest effect size detected across models), above and beyond the effect of academic stressful events. Social stressful events were also accompanied by reduced time spent on schoolwork, perhaps reflecting general disruptions in students' ability or desire to concentrate on work. This might reflect that social stressful events are perceived as more personally meaningful and uniquely challenging than academic stressful events, especially for students undergoing the transition to college and its associated restructuring of social relationships and networks (Paul & Brier, 2001; Rose, 1984; Swenson et al., 2008).

Although an elevation in feelings of sadness and anxiety appear to be a normative proximate response following interpersonal stressful events, repeated interpersonal tensions and conflict may be a marker of more enduring symptoms of psychological distress. Exploratory correlational analyses found that participants who reported a higher rate of social stressful events, as well as those with a higher proportion of negative mentions of other people in their daily diaries, also scored higher on a clinical inventory of global symptoms of psychopathology. These results do not establish causality, and it is possible that students with higher baseline symptoms of psychopathology are more likely to encounter social conflict due to their already elevated negative affect (Daley et al., 1997; Hammen, 1991; McLaughlin & Nolen-Hoeksema, 2012; Rudolph et al., 2009). Further longitudinal research should evaluate the prospective role of daily social stressful event exposure in the emergence and progression of longer-term symptoms of psychopathology, including whether this role differs significantly from the contributions of other forms of life stress (Brown & Harris, 1978; O'Neill et al., 2004; Parker et al., 2015; Rudolph et al., 2009, 2016; Whisman & Beach, 2010).

The present results should be considered in light of their methodological limitations. First, our "academic" and "social" life stress categories require further refinement in future

research, including at the individual level. For example, although both pertaining to the academic life domain, heightened workload associated with preparing for an exam and receiving a bad grade are likely to elicit different sets of behaviors and affective responses. Moreover, a strict separation between academic and social domains might be difficult in some cases; e.g., getting a bad grade could be argued to have social significance, as it involves an evaluation of the student's performance by someone else and might trigger peer comparisons. Relatedly, additional features of a stressful event might be more indicative of the event's potential impact on students' behavior and wellbeing. For example, stress researchers have referred to different "domains of threat" to characterize a stressful event's subjective meaning to an individual, including "loss", "danger", and "disappointment" (Almeida & Horn, 2004; Lazarus & Folkman, 1984). Future research could use similar categorizations, in conjunction with life domain, to provide a finer-grained characterization of the features of stressful events that might be most closely related to their impact on behavior, affect, and mental health outcomes.

Second, our identification of daily stressful events relied on participants' self-recorded daily voice diaries, but we had missing data and participants may have left out important information relevant to our analyses. All participants shared at least some instances of both academic and social stressful events, but it is possible that participants did not always provide a comprehensive account. Third, our naturalistic intensive longitudinal design lends ecological validity to the examination of life stress, but our study was observational and our models testing the associations among stress, affect, and behavior did not ascertain causality. Finally, although we tested both same-day and next-day effects, we did not assess the possibility of nonlinear or cumulative effects of daily stressful events over time.

Conclusion

The present study leveraged mobile and wearable technology and cutting-edge natural language processing techniques to examine the affective and behavioral signatures of different domains of life stress in daily life. Our results demonstrate the utility of deep phenotyping approaches for the investigation of real-world life stress and suggests that academic and social dimensions of life stress may have distinct effects on daily affect and behavior, with potential implications for long-term wellbeing.

Acknowledgments

Laura Farfel, Marisa Marotta, Erin Phlegar, Lauren DiNicola, and Arpi Youssoufian helped collecting data. Timothy O'Keefe, Harris Hoke, and Lily Jeong provided informatics assistance. Kenzie W. Carlson provided support with Beiwe. Katherine Miclau, Amira Song, Emily Iannazzi and Hannah Becker helped with actigraphy data quality control. Jennie Li helped with voice diary data quality control. Katie McLaughlin provided advice on our adaptation of the Life Stress Interview scoring for the identification of stressful events in daily diaries, and Isabelle Jacques, Adelaide Kelsey, Marina Bibars, and Olivia Bolitho supported the execution of this scoring system. This work was supported by a generous gift from Kent and Liz Dauten, National Institute of Mental Health grants U01MH116925 and DP2MH103909, National Institutes of Health grant T90DA022759, Harvard University's Rand Fund for Innovation in Psychology Research, and Harvard University's Sackler Scholar Program in Psychobiology.

Data Availability

The data sets generated during this study are not publicly available due to concerns related to participant identifiability but may be available from the corresponding author on reasonable request.

Authors' Contributions

G.C., J.P.O., J.T.B., and R.L.B. designed the study. G.C. and R.L.B. collected the data. C.V.B, H.R.E., and R.L.B. analyzed the data. P.M. provided statistical support. C.V.B. and R.L.B. wrote the paper, and all other authors reviewed the final manuscript.

Competing Interests

J.P.O is a cofounder and board member of Phebe, a commercial entity that operates in digital phenotyping. J.T.B has received consulting fees from Verily Life Sciences as well as consulting fees and equity from Mindstrong Health Inc. for work unrelated to the present work. R.L.B. has received consulting fees from Pfizer, Roche, Alkermes, and Cognito for work unrelated to the present work. All other authors declare no competing interests.

References

Almeida, D. M. (2005). Resilience and Vulnerability to Daily Stressors Assessed via Diary Methods. *Current Directions in Psychological Science*, *14*(2), 64–68. https://doi.org/10.1111/j.0963-7214.2005.00336.x

Almeida, D. M., & Horn, M. (2004). Is Daily Life More Stressful During Middle Adulthood? In *How Healthy Are We?: A National Study of Well-Being at Midlife.* (pp. 425–451). The University of Chicago Press.

Almeida, D. M., Wethington, E., & Kessler, R. (2002). The Daily Inventory of Stressful Events: An Interview-Based Approach for Measuring Daily Stressors. *Assessment*, *9*(1), 41–55. https://doi.org/10.1177/1073191102091006

American College Health Association. (2023). *American College Health Association-National College Health Assessment III: Undergraduate Student Reference Group Data Report Spring 2023*. American College Health Association.

Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M., & Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders*, *173*. https://doi.org/10.1016/j.jad.2014.10.054

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Bolger, N., DeLongis, A., Kessler, R. C., & Schilling, E. A. (1989). Effects of daily stress on negative mood. *Journal of Personality and Social Psychology*, *57*(5), 808–818. https://doi.org/10.1037/0022-3514.57.5.808

Brown, G., & Harris, T. (1978). *Social Origins of Depression: A study of psychiatric disorder in women*. Routledge. https://www.routledge.com/Social-Origins-of-Depression-A-study-of-psychiatric-disorder-in-women/Brown-Harris/p/book/9780415510929

Cacioppo, J. T., & Hawkley, L. C. (2003). Social Isolation and Health, with an Emphasis on Underlying Mechanisms. *Perspectives in Biology and Medicine*, *46*(3), S39–S52.

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu

(Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Charles, S. T., Piazza, J. R., Mogle, J., Sliwinski, M. J., & Almeida, D. M. (2013). The Wear and Tear of Daily Stressors on Mental Health. *Psychological Science*, *24*(5), 733–741. https://doi.org/10.1177/0956797612462222

Cohen, S. (2004). Social Relationships and Health. *American Psychologist*, *59*(8), 676–684. https://doi.org/10.1037/0003-066X.59.8.676

Daley, S. E., Hammen, C., Burge, D., Davila, J., Paley, B., Lindberg, N., & Herzberg, D. S. (1997). Predictors of the generation of episodic stress: A longitudinal study of late adolescent women. *Journal of Abnormal Psychology*, *106*(2), 251–259. https://doi.org/10.1037/0021-843X.106.2.251

daSilva, A. W., Huckins, J. F., Wang, W., Wang, R., Campbell, A. T., & Meyer, M. L. (2021). Daily perceived stress predicts less next day social interaction: Evidence from a naturalistic mobile sensing study. *Emotion*, *21*(8), 1760–1770. https://doi.org/10.1037/emo0000994

DeLongis, A., & Holtzman, S. (2005). Coping in Context: The Role of Stress, Social Support, and Personality in Coping. *Journal of Personality*, *73*(6), 1633–1656. https://doi.org/10.1111/j.1467-6494.2005.00361.x

Derogatis, L. R. (1975). *SCL-90-R: Symptom Checklist-90-R : Administration, Scoring, and Procedures Manual*. NCS Pearson. https://books.google.com/books/about/SCL_90_R.html?id=fcxxtwAACAAJ

Derogatis, L. R., & Unger, R. (2010). Symptom Checklist-90-Revised. In *The Corsini Encyclopedia of Psychology* (pp. 1–2). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216.corpsy0970

Dyson, R., & Renk, K. (2006). Freshmen adaptation to university life: Depressive symptoms, stress, and coping. *Journal of Clinical Psychology*, *62*(10). https://doi.org/10.1002/jclp.20295

Folkman, S., & Moskowitz, J. T. (2004). Coping: Pitfalls and Promise. *Annual Review of Psychology*, *55*(1), 745–774. https://doi.org/10.1146/annurev.psych.55.090902.141456

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2023). *rstanarm: Bayesian applied regression modeling via Stan* (2.21.4) [Computer software]. https://mc-stan.org/rstanarm/

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Hammen, C. (1988). Self-cognitions, stressful events, and the prediction of depression in children of depressed mothers. *Journal of Abnormal Child Psychology*, *16*(3), 347–360. https://doi.org/10.1007/BF00913805

Hammen, C. (1991). Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology*, *100*(4), 555–561.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. *Perspectives on Psychological Science*, *11*(6). https://doi.org/10.1177/1745691616650285

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Holi, M. M., Sammallahti, P. R., & Aalberg, V. A. (1998). A Finnish validation study of the SCL-90. *Acta Psychiatrica Scandinavica*, *97*(1), 42–46. https://doi.org/10.1111/j.1600-0447.1998.tb09961.x

Kay, M. (2021). *tidybayes: Tidy Data and Geoms for Bayesian Models* (3.0.1) [Computer software]. http://mjskay.github.io/tidybayes/

Kim, E.-J., & Dimsdale, J. E. (2007). The Effect of Psychosocial Stress on Sleep: A Review of Polysomnographic Evidence. *Behavioral Sleep Medicine*, *5*(4), 256–278. https://doi.org/10.1080/15402000701557383

Lazarus, R. (1999). *Stress and emotion: A new synthesis* (pp. xiv, 342). Springer Publishing Co.

Lazarus, R., & Folkman, S. (1984). *Stress, Appraisal, and Coping*. Springer Publishing Company.

Lee-Baggley, D., Preece, M., & DeLongis, A. (2005). Coping With Interpersonal Stress: Role of Big Five Traits. *Journal of Personality*, *73*(5), 1141–1180. https://doi.org/10.1111/j.1467-6494.2005.00345.x

Lipson, S. K., Zhou, S., Abelson, S., Heinze, J., Jirsa, M., Morigney, J., Patterson, A., Singh, M., & Eisenberg, D. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *Journal of Affective Disorders*, *306*, 138–147. https://doi.org/10.1016/j.jad.2022.03.038

Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., & Camacho-Collados, J. (2022, February 8). *TimeLMs: Diachronic Language Models from Twitter*. arXiv.Org. https://arxiv.org/abs/2202.03829v2

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. https://doi.org/10.48550/arXiv.1802.03426

McLaughlin, K. A., & Nolen-Hoeksema, S. (2012). Interpersonal stress generation as a mechanism linking rumination to internalizing symptoms in early adolescents. *Journal of Clinical Child and Adolescent Psychology*, *41*(5). https://doi.org/10.1080/15374416.2012.704840

Morey, R., & Rouder, J. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs* (0.9.12-4.7) [Computer software]. https://CRAN.R-project.org/package=BayesFactor

Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology*, *14*(2), 99–119. https://doi.org/10.20982/tqmp.14.2.p099

O'Neill, S. C., Cohen, L. H., Tolpin, L. H., & Cimbolic Gunthert, K. (2004). Affective reactivity to daily interpersonal stressors as a prospective predictor of depressive symptoms. *Journal of Social and Clinical Psychology*, *23*(2). https://doi.org/10.1521/jscp.23.2.172.31015

Onnela, J.-P., Dixon, C., Griffin, K., Jaenicke, T., Minowada, L., Esterkin, S., Siu, A., Zagorsky, J., & Jones, E. (2021). Beiwe: A data collection platform for high-throughput digital phenotyping. *Journal of Open Source Software*, *6*(68), 3417. https://doi.org/10.21105/joss.03417

Parker, J. G., Rubin, K. H., Erath, S. A., Wojslawowicz, J. C., & Buskirk, A. A. (2015). Peer Relationships, Child Development, and Adjustment: A Developmental Psychopathology Perspective. In *Developmental Psychopathology* (pp. 419–493). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470939383.ch12

Paul, E. L., & Brier, S. (2001). Friendsickness in the Transition to College: Precollege Predictors and College Adjustment Correlates. *Journal of Counseling & Development*, *79*(1), 77–89. https://doi.org/10.1002/j.1556-6676.2001.tb01946.x

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rahimi-Eichi, H., Coombs 3rd, G., Vidal Bustamante, C. M., Onnela, J.-P., Baker, J. T., & Buckner, R. L. (2021). Open-source Longitudinal Sleep Analysis From Accelerometer Data (DPSleep): Algorithm Development and Validation. *JMIR MHealth and UHealth*, *9*(10). https://doi.org/10.2196/29849

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Rose, S. (1984). How Friendships End: Patterns among Young Adults. *Journal of Social and Personal Relationships*, *1*(3). https://doi.org/10.1177/0265407584013001

Rosenthal, S., Farra, N., & Nakov, P. (2019). *SemEval-2017 Task 4: Sentiment Analysis in Twitter* (arXiv:1912.00741). arXiv. https://doi.org/10.48550/arXiv.1912.00741

Ross, S., Niebling, B., & Heckert, T. (1999). Sources of Stress among College Students. *College Student Journal*, *33*(2).

Rudolph, K. D., Flynn, M., Abaied, J. L., Groot, A., & Thompson, R. (2009). Why is Past Depression the Best Predictor of Future Depression? Stress Generation as a Mechanism

of Depression Continuity in Girls. *Journal of Clinical Child & Adolescent Psychology*, *38*(4), 473–485. https://doi.org/10.1080/15374410902976296

Rudolph, K. D., Lansford, J. E., & Rodkin, P. C. (2016). Interpersonal Theories of Developmental Psychopathology. In *Developmental Psychopathology*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119125556.devpsy307

Rytilä-Manninen, M., Fröjd, S., Haravuori, H., Lindberg, N., Marttunen, M., Kettunen, K., & Therman, S. (2016). Psychometric properties of the Symptom Checklist-90 in adolescent psychiatric inpatients and age- and gender-matched community youth. *Child and Adolescent Psychiatry and Mental Health*, *10*(1), 23. https://doi.org/10.1186/s13034-016-0111-x

Sadeh, A., & Gruber, R. (2002). Stress and Sleep in Adolescence: A Clinical-Developmental Perspective. In M. A. Carskadon (Ed.), *Adolescent Sleep Patterns: Biological, social, and psychological influences* (pp. 236–253). Cambridge University Press. https://doi.org/10.1017/CBO9780511499999.017

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., & Picard, R. (2018). Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. *Journal of Medical Internet Research*, *20*(6), e210. https://doi.org/10.2196/jmir.9410

Schmitz, N., Hartkamp, N., Kiuse, J., Franke, G. H., Reister, G., & Tress, W. (2000). The Symptom Check-List-90-R (SCL-90-R): A German validation study. *Quality of Life Research*, *9*(2), 185–193. https://doi.org/10.1023/A:1008931926181

Smyth, J. M., Zawadzki, M., & Gerin, W. (2013). Stress and Disease: A Structural and Functional Analysis. *Social and Personality Psychology Compass*, *7*(4), 217–227. https://doi.org/10.1111/spc3.12020

Smyth, J. M., Zawadzki, M. J., Marcusson-Clavertz, D., Scott, S. B., Johnson, J. A., Kim, J., Toledo, M. J., Stawski, R. S., Sliwinski, M. J., & Almeida, D. M. (2023). Computing Components of Everyday Stress Responses: Exploring Conceptual Challenges and New Opportunities. *Perspectives on Psychological Science*, *18*(1), 110–124. https://doi.org/10.1177/17456916221082108

Smyth, Sliwinski, M. J., Zawadzki, M. J., Scott, S. B., Conroy, D. E., Lanza, S. T., Marcusson-Clavertz, D., Kim, J., Stawski, R. S., Stoney, C. M., Buxton, O. M., Sciamanna, C. N., Green, P. M., & Almeida, D. M. (2018). Everyday stress response targets in the science of behavior change. *Behaviour Research and Therapy*, *101*, 20–29. https://doi.org/10.1016/j.brat.2017.09.009

Stan Development Team. (2018). *The Stan Core Library* [Computer software]. http://mc-stan.org

Stults-Kolehmainen, M. A., & Sinha, R. (2014). The Effects of Stress on Physical Activity and Exercise. *Sports Medicine*, *44*(1), 81–121. https://doi.org/10.1007/s40279-013-0090-5

Swenson, L. M., Nordstrom, A., & Hiester, M. (2008). The Role of Peer Relationships in Adjustment to College. *Journal of College Student Development*, *49*(6), 551–567.

Tennen, H., Suls, J., & Affleck, G. (1991). Personality and Daily Experience: The Promise and the Challenge. *Journal of Personality*, *59*(3), 313–337. https://doi.org/10.1111/j.1467-6494.1991.tb02387.x

Teychenne, M., Costigan, S. A., & Parker, K. (2015). The association between sedentary behaviour and risk of anxiety: A systematic review. *BMC Public Health*, *15*(1), 513. https://doi.org/10.1186/s12889-015-1843-x

Thorley, C. (2017). *Not By Degrees: Improving student mental health in the UK's Universities*. Institute for Public Policy Research.

Tomiyama, A. J. (2019). Stress and Obesity. *Annual Review of Psychology*, *70*(1), 703–718. https://doi.org/10.1146/annurev-psych-010418-102936

Van Reeth, O., Weibel, L., Spiegel, K., Leproult, R., Dugovic, C., & Maccari, S. (2000). Interactions between stress and sleep: From basic research to clinical situations. *Sleep Medicine Reviews*, *4*(2), 201–219. https://doi.org/10.1053/smrv.1999.0097

Vidal Bustamante, C. M., Coombs, G., Rahimi-Eichi, H., Mair, P., Onnela, J.-P., Baker, J. T., & Buckner, R. L. (2022). Fluctuations in behavior and affect in college students measured using deep phenotyping. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-05331-7

Vidal Bustamante, C. M., Coombs III, G., Rahimi-Eichi, H., Mair, P., Onnela, J.-P., Baker, J. T., & Buckner, R. L. (2024). Precision Assessment of Real-World Associations Between Stress and Sleep Duration Using Actigraphy Data Collected Continuously for an

Academic Year: Individual-Level Modeling Study. *JMIR Formative Research*, *8*(1), e53441. https://doi.org/10.2196/53441

Vidal Bustamante, C. M., Rodman, A. M., Dennison, M. J., Flournoy, J. C., Mair, P., & McLaughlin, K. A. (2020). Within-person fluctuations in stressful life events, sleep, and anxiety and depression symptoms during adolescence: A multiwave prospective study. *Journal of Child Psychology and Psychiatry*, *61*(10), 1116–1125. https://doi.org/10.1111/jcpp.13234

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3–14. https://doi.org/10.1145/2632048.2632054

Whisman, M. A., & Beach, S. R. H. (2010). Models for understanding interpersonal processes and relationships in anxiety disorders. In *Interpersonal processes in the anxiety disorders: Implications for understanding psychopathology and treatment* (pp. 9–35). American Psychological Association. https://doi.org/10.1037/12084-001

Zautra, A. J. (2006). *Emotions, Stress, and Health*. Oxford University Press.

Zhang, J., Paksarian, D., Lamers, F., Hickie, I. B., He, J., & Merikangas, K. R. (2017). Sleep Patterns and Mental Health Correlates in US Adolescents. *The Journal of Pediatrics*, *182*, 137–143. https://doi.org/10.1016/J.JPEDS.2016.11.007

Supplementary Materials to Accompany "Year-Long Digital Phenotyping and Natural Language Processing of Daily Voice Diaries Reveal Affective and Behavioral Signatures of Real-World Life Stress"

Constanza M. Vidal Bustamante[a,b], Garth Coombs 3rd[a,b], Habiballah Rahimi-Eichi[a-d], Patrick Mair[a], Jukka-Pekka Onnela[e], Justin T. Baker[c,d], and Randy L. Buckner[a-c,f]

[a] Department of Psychology, Harvard University, Cambridge, Massachusetts, USA
[b] Center for Brain Science, Harvard University, Cambridge, Massachusetts, USA
[c] Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA
[d] Institute for Technology in Psychiatry, McLean Hospital, Belmont, Massachusetts, USA
[e] Department of Biostatistics, Harvard University, Boston, Massachusetts, USA
[f] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, Charlestown, Massachusetts, USA

Correspondence should be addressed to Constanza M. Vidal Bustamante at cvidal@g.harvard.edu

Contents

Supplementary Methods: Daily Phone-Based Survey Questions Used in Analysis

How much of your waking time did you spend studying or doing homework over the past 24 hours?
1) Very little of my time (0-20%)
2) Some of my time (21-40%)
3) About half of my time (41-60%)
4) Most of my time (61-80%)
5) Almost all of my time (81-100%)

How much of your waking time did you spend interacting with others over the past 24 hours?
1) Very little of my time (0-20%)
2) Some of my time (21-40%)
3) About half of my time (41-60%)
4) Most of my time (61-80%)
5) Almost all of my time (81-100%)

How much did you want to spend time with others in the past 24 hours?
-5 = Wanted to be completely alone
0 = Neutral
5 = Wanted to be completely with others

-5	-4	-3	-2	-1	0	1	2	3	4	5

How much did you feel nervous or anxious over the past 24 hours?
1) Very slightly or not at all
2) A little
3) Moderately
4) Quite a bit
5) Extremely

How much did you feel upset over the past 24 hours?
1) Very slightly or not at all
2) A little
3) Moderately
4) Quite a bit
5) Extremely

How much did you feel irritable over the past 24 hours?
1) Very slightly or not at all

2) A little
3) Moderately
4) Quite a bit
5) Extremely

How much did you feel angry or hostile over the past 24 hours?
1) Very slightly or not at all
2) A little
3) Moderately
4) Quite a bit
5) Extremely

How much did you feel sad or down over the past 24 hours?
1) Very slightly or not at all
2) A little
3) Moderately
4) Quite a bit
5) Extremely

Supplementary Methods: Voice Diary Instructions to Participants

1. Please complete your daily voice recording directly after completing the survey questions.
2. Please describe how you felt over the last 24 hours in relation to events that have occurred as well as to upcoming events either specific or general that are on your mind.
3. Please speak for at least 1 minute (but it will automatically stop after 4 minutes).
4. Please do talk about important people and events that might be encountered or have been encountered, but avoid specific details that could identify those people or locations. For example, you could say "I was feeling lonely because I spent all day in my dorm" or "I was in a good mood because I'm excited about a nice restaurant I'm going to tomorrow," but please don't say "I'm excited about going to Union Oyster House tomorrow" or "I spent all day in the Lowell House dorm." It would be fine to say "I've been feeling anxious about seeing a family member that I tend to fight with during the holidays," but not "I've been feeling anxious about seeing my brother Tom this coming holiday."
5. Give a good sense of how you've been feeling that day and why and about specific upcoming events that are on your mind: if you're nervous about an upcoming assignment, got into a fight with someone, excited about good things that have happened or that you're looking forward to. For instance, some things you could talk about would be (but not limited to):
    a. Who you interact with today (remember, no names)? How close were these people to you (closest friends, acquaintances, etc.)? How do you feel the interactions went?
    b. Did you feel particularly accepted, rejected or emotionally supported and why?
    c. What were the highs and lows of your day? Did anything particularly notable happen, and how did it make you feel?
    d. What were the things that made you particularly stressed? Relaxed? Happy? Sad? How did you deal with them?
6. When you are ready, press the "Done" button, which will submit your recording.

Supplementary Methods: Health Questionnaire

Please specify which of the following conditions you have now or have had in the past.

	In Your Lifetime	Currently (Now)	N/A
Cancer (other than skin cancer)			
Chronic migraine headaches			
Diabetes			
Encephalitis or meningitis			
Epilepsy			
High blood pressure			
Multiple sclerosis			
Parkinson's disease			
Stroke			
Depression (diagnosed or treated)			
Attention Deficit Hyperactivity Disorder			
Anxiety disorder			
Personality disorder			
Schizophrenia or other psychotic disorder			
Bipolar disorder			
Alcohol Abuse or Dependence			
Other Substance Abuse or Dependence			
Other significant illnesses			
Please specify			

Supplementary Methods: Scoring System Used to Identify Stressful Events from Participants' Daily Voice Diaries

Rater Training Procedure

Training the team of raters involved three stages. First, the master rater went over each of the diaries for one of the pilot participants and explained to the raters the correct way to score each diary. Secondly, for an additional two pilot participants' data, the master rater went over the first fifteen diaries together with the raters, and raters were then asked to score that participant's remaining diaries by themselves, one third of diaries at a time. After they were done with one third, the master rater went over each of the diaries together with the rater and any discrepancies were discussed and corrected to match the master rater's answer. The same procedure was repeated for the remaining two thirds of that participant's data, and then the process was repeated for another pilot participant. Lastly, for the final pilot participant, the master rater went over the first fifteen diaries together with the rater, and then the rater scored all remaining diaries by themselves. The master rater then computed the overall accuracy of each rater's answers (compared to the master rater's answers); if the accuracy was 90% or higher, the rater was deemed ready for the independent scoring of target participants' data. All four raters passed this bar (and all incorrect answers were discussed).

Written Summary of Instructions

Below is a written summary of instructions provided to the team of trained raters.

General Instructions

We will read each transcript of participants' daily diaries and identify and mark stressful events mentioned by the participant according to the following attributes:

- Life domain: Academic OR Social OR Other (Health, Finances, Other)
- Severity: Moderate OR Severe

We will use a spreadsheet to keep track of our scores. Each spreadsheet will have a single subject's data, with each row representing a diary transcript. Each column to the right will be used to note events belonging to the combination of categories mentioned above (e.g., Academic Severe, Social Moderate, etc.).

1. Start by reading the transcript
2. For any stressful event identified in the transcript, pick the appropriate column/category to which it belongs

3. Write a brief note describing the event on that cell

Additional Notes

Score all diaries for a given subject in one sitting

- When assessing the intensity of the daily events, we are primarily concerned with within-participant comparisons, i.e., events that are relatively more (or less) intensely positive/negative for that participant. Therefore, we will score all of a participant's diaries close in time so you remember what is typical/unusual for each subject.

Consult the repository of examples/notes when uncertain

- Sometimes you might be unsure about how to score some of the diaries. Please consult these notes and examples of stressful events at the end of this document frequently and every time you are unsure about whether/how to score a particular segment of the diary.
- Sometimes you might be unsure whether a certain event belongs in the Academic or Social life domain. For example, a participant might mention a tense interaction with classmates during class. In such cases, focus on the specific source of stress. In the earlier example, it is the interaction that is stressful, so it would belong to the Social domain. If the source of stress is mostly related to the participant's academic performance, then it would belong to the Academic domain.
- If after consulting these notes you are still unsure about how to score a diary, leave a note about it under the "Notes" column of that participant's spreadsheet and make sure to check with the master rater during the mid-point and final check-ins.

Scoring past / upcoming events

- We are primarily concerned with scoring any ***events* that happened *the day of the transcript***
- Do not mark an event the day it was mentioned if:
    - The participant mentions *thinking* about something that happened in the past, but no event (or related event) happened the day the diary was submitted. In these cases, you can add a note about it under "Notes" (e.g. "thinks about / feels sad about person they broke up with recently")
    - The participant mentions *missing* home / family / friends
    - The participant mentions being *worried* about upcoming exams / assignments / etc.
    - In the three cases above, these are just thoughts/feelings that are not themselves "events". That said, if there is a specific event that triggered the thoughts/feelings, think if that event itself could be considered under one of the event categories in the system.
- Do mark as event the day it was mentioned if:

- The participant mentions making plans *today* for an exciting event in the future
- The participant mentions heightened workload *today* related to preparing for event in the future (e.g., working long hours on a submission due the next day)
- In the two cases above, the acts of planning/preparing are themselves events that we are noting

Examples of Academic and Social Stressful Events to note			
ACADEMIC STRESSFUL EVENTS		SOCIAL STRESSFUL EVENTS	
Severe	Moderate	Severe	Moderate
Consequential bad / much worse-than-expected grade / performance	Worse than expected grade / performance in assignment/activities, but not terrible or too consequential (e.g., on small quiz)	Argument / confrontation with others	Hassles with close others, e.g. friends/roommates acting weird, annoying
Extremely demanding workload, overwhelmed / struggling to complete on time	Demanding workload, difficult to complete, unable to focus	Break-ups or strains in relationships (romantic or otherwise)	Uncomfortable / awkward / disappointing social interactions with others (in absence of confrontation)
		Hearing very bad news about loved ones	Do NOT mark transient annoyances with people who are not part of the participant's life (e.g., someone in the dining hall being loud)

Supplementary Table 1: Available Daily Observations by Participant

Supplementary Table 1. Available daily observations by participant and data stream				
Participant	Diaries	Surveys	Actigraphy: Sleep Duration	Actigraphy: Waketime Activity
2BK68	89	150	205	201
2EG95	217	218	216	214
2FB35	146	147	194	194
2XQ45	98	112	203	202
3BY45	162	198	198	197
3EF35	201	204	205	206
3HZ25	157	171	163	163
3JT35	124	182	204	203
3KN45	107	116	184	185
3KS75	214	217	210	210
3MR58	158	185	197	197
3PK35	70	117	165	165
3RV48	221	223	214	213
4EF58	87	93	204	203
4NB88	211	213	103	103
4SX38	166	171	185	185
4WH78	199	204	208	207
4XP98	99	104	185	184
5EX58	206	206	202	202
5PT78	209	211	192	192
5SD58	211	212	211	211
5TH28	128	131	158	159
5VT38	83	88	211	211
5ZK55	164	172	177	177
6BN38	204	206	197	196
6CQ38	218	223	205	205
6NH38	184	185	195	195
6QF48	218	218	214	214
6VQ88	219	219	198	199
6YG55	206	210	206	205
7EK65	212	214	209	208
7MF65	130	136	194	194

7VF65	107	123	194	193
7XP88	208	211	216	216
7ZW55	218	216	212	211
8AU38	191	200	206	205
8EY75	204	206	184	183
8GA95	160	149	204	203
8GW68	171	174	115	114
8MB75	107	127	163	165
8NZ48	84	160	174	174
8QZ55	216	216	199	199
8RC95	180	176	210	209
8TS58	177	187	203	199
8VY68	205	205	208	207
9AR68	211	212	204	203
9RP78	191	193	194	193
9SA35	206	205	197	197
9TQ38	126	191	195	195

Supplementary Table 2: Topic Modeling with BERTopic

Supplementary Table 2. Results of topic modeling of students' daily diaries with BERTopic.					
Cluster Size Rank	**Top 20 Representative Words**	**Sample Sentences**	**# Sentences**	**Topic Label**	**Category**
0	class - dinner - homework - went - lunch - classes - then - to - had - and - my - breakfast - um - the - for - uh - of - after - studying - some	['Um, and then went back, got dinner, um, came back to the dorm, did some work.', 'And then I went to class and then lunch and it was all pretty uneventful.', 'Um, I had class in the morning 10:00 to 11:30 and then class in the afternoon from 1:30, um-- from 1:00 right through till 4:00, 4:30, um, which is a lot more time than I spent in class for most days last semester.']	16086	Daily schedule	schedule
19	back - came - room - then - home - work - after - we - went - lost - and - won - here - headed - um - to - got - uh - did - my	['And then, I came back to my room.', 'Um, and then I came back to my room for a bit.', 'Um, then we came back, and I s-- largely spent my time alone, just, like, doing work.']	796	Got back	schedule

64	didn - much - do - anything - as - really - would - liked - whole - did - have - else - not - lot - that - but - other - what - um - two	["Um, didn't really do much.", "Uh, didn't really do a whole lot.", "I didn't really do anything at all."]	214	Didn't do much	schedule
85	hours - minutes - three - hour - two - five - seven - six - half - an - four - like - almost - long - it - 15 - worked - maybe - 45 - for	["I spent like six, seven hours with these two just constantly.", "So that's like three hours.", "That 15 minutes is all the difference in the world."]	125	Duration	schedule
17	tomorrow - work - do - ll - hopefully - have - be - done - get - gonna - will - lot - looking - to - day - stuff - so - more - forward - should	[' So I have a lot of work to do tomorrow and, um, I spent today working on that stuff, but I still have more to do tomorrow morning, um, to finish it off.', 'And not too much going on tomorrow.', 'Uh, don't know what I have coming up tomorrow.]	966	Tomorrow	schedule
9	week - weekend - this - thanksgiving - be - friday - break - looking	['This is gonna be a good week.', "This week's really just gonna be-- it's gonna be	1854	Week	schedule

	- saturday - sunday - busy - forward - next - spring - thursday - gonna - have - schedule - is - wednesday	a week.", 'Um, I do have a lot of work this weekend and, like, this coming week.']			
3	day - was - today - good - it - pretty - fun - overall - yeah - interesting - so - weird - really - um - great - okay - that - been - but - bad	['Um, today was a really good day.', ' So I didn't really have a great day today.', ' Today has been a pretty typical day.']	7887	My day	feelings
77	night - good - tonight - evening - fun - last - was - great - it - yeah - overall - really - pretty - another - but - ended - so - an - nice - um	['Uh, tonight was kind of an unplanned late night.', ' Had a really decent night.', 'Uh, but nonetheless, a pretty fun last night with the boys.']	159	My night	feelings
46	going - are - things - well - everything - fine - okay - pretty - yeah - life - is - right - think - general - than - all - um - overall - uh - other	['Things are going well.', 'But I mean, I guess it's okay because I think I have everything done.', 'Um, yeah, I mean things are going pretty well I would say, nothing really too	312	Going well	feelings

		drastically bad or good.']			
49	bad - not - good - great - terrible - it - thing - crazy - too - yeah - that - pretty - clutch - nothing - really - so - looks - awful - nice - bro	['It could be better, but it's not terrible.'", "But it's not that bad.", "Not, not really that bad, so, you know, that's good."]	297	Bad	feelings
74	did - well - think - mistakes - questions - wrong - mistake - messed - don - stupid - but - it - one - there - didn - made - knew - really - bad - pretty	['I think I did really well.', 'I know that there are multiple things that I did wrong or could have done better, so that is frustrating.', 'And I, I guess I did well enough to make the roster, so that was good news.']	172	Performance	feelings
4	happy - feel - feeling - felt - mood - sad - good - about - really - it - just - pretty - like - so - that - better - very - me - but - was	['Um, and I felt really happy in those moments.', "And I was actually pretty pleased with how it went.", 'And I think about that a lot, and that kinda, like, put me in a bad mood, like a sad	2638	Happy/Unhappy	feelings

		mood for a while, which is weird to say, but whatever.']			
7	stressed - stressful - stress - stressing - out - about - not - little - of - just - that - but - me - so - like - it - bit - very - kind - because	["Um, so that's definitely a source of some stress right now.", "Right now, I'm more stressed than anything else, really.", "I don't know that I'm super stressed about it."]	2256	Stress	feelings
47	mind - thinking - lot - think - on - much - about - thoughts - been - just - too - not - that - going - it - yeah - there - else - trying - what	['But there's just something on my mind that just keeps bringing me back.', 'Um, yeah, so just a lot of-- a lot of stuff on my mind.', 'And that is really taking precedent in my mind.']	310	On my mind	feelings
14	myself - like - feel - you - self - me - proud - not - just - don - people - know - be - that - want - in - of - life - it - hate	['So I'm disappointed in myself that I'm not-- that I didn't have a stronger start here when I could have.', 'But I will be kind to myself and not be hard on myself, uh, for not being that way.', "But I don't know what	1195	Self reflection	feelings

		to do with myself."]			
18	tired - exhausted - energy - just - very - really - super - now - felt - was - so - feel - sleepy - right - yeah - like - feeling - but - throughout - um	["Uh, and I'm mentally exhausted, but not physically at all because I didn't move physically.", 'Um, it just feels like I'm, I'm tired all the time, even though I'm sleeping like, eight, eight and a half, nine hours, taking naps and stuff.', 'Um, had-- my energy was pretty consistently high throughout the entire day.']	944	Tiredness	feelings
28	relaxed - relaxing - relax - day - pretty - calm - just - very - was - today - it - bit - nice - relaxation - more - um - and - been - really - down	[' Um, so today was a pretty relaxed day.', 'Um, but it was, like, a pretty relaxed day.', ' So today was a pretty good day, um, pretty relaxed-- pretty, um, relaxing.']	541	Relaxation	feelings

30	nervous - anxiety - anxious - about - little - like - panic - feeling - but - just - bit - really - of - think - been - that - ve - attack - it - because	['And I'm, I'm really nervous that I'm just gonna have to know a lot of material.', 'Um, I was anxious in the morning about some things I said to a friend last night, um, 'cause I was worried that she, like, thought less of me because of them, um, but now I think that that's fine.', 'Um, I am very nervous 'cause I was just reading that I have to do-- but I don't understand it.']	503	Anxiety	feelings
35	excited - exciting - super - hype - very - hyped - about - so - really - that - for - yeah - it - pretty - um - be - was - but - excitement - pumped	['And I'm excited to see what these opportunities bring.', 'Um, I feel excited about what I'm doing.', 'But I'm not that excited for it as I should be.']	436	Excitement	feelings
36	chill - chilling - chilled - day - pretty - just - been - kinda - was - yeah - it - today - be - out	['Um, and then, you know, just a very mellow, chill day.', 'It was chill.', 'That was a	416	Chill	feelings

	- very - uh - like - really - so - um	good way to just chill out a bit.']			
43	focus - distracted - concentrate - focused - couldn - focusing - can - could - just - hard - on - distracting - really - work - not - trouble - kept - able - like - was	["Um, I just couldn't focus on it.", "Um, then I tried to do work, couldn't focus, couldn't concentrate.", "But I really couldn't focus on it."]	334	Focus	feelings
55	worried - worry - about - scared - too - not - concerned - worrying - that - it - something - little - be - don - gonna - very - but - so - much - really	["But there's something that's nagging me that's saying that I should be a little worried.", "I'm worried that things are gonna fall apart 'cause that seems like something that might happen and I'm gonna, you know, knock on wood.", "Yeah, it's not really something to be really that worried about."]	235	Worry	feelings

58	lonely - alone - isolated - feel - feeling - miss - felt - just - kind - very - missed - like - myself - them - little - didn - they - ve - of - me	['I kinda felt alone in that sense, um, and not really having anyone to talk to or anyone that fully understands me and, like, what I'm going through.', 'And I was just kind of overcome by the fact that I was, like, really alone and wanted to be with someone.', "I don't know, I guess I'm feeling a little lonely even though I'm surrounded by people."]	226	Loneliness	feelings
73	hard - difficult - it - easy - harder - over - tough - just - really - so - not - tricky - is - don - shouldn - but - struggle - that - easier - getting	['I just found it really difficult, just real tough day , so yeah.', 'Um, so it's definitely been a struggle.', 'There's just, like, a lot going on right now, and I don't really know how to handle everything.']	180	Hard	feelings
81	frustrating - understand - concepts - understood - kind - it - that -	['So that was kind of frustrating.', 'And that would happen like every time, um, four	137	Frustrating	feelings

	they - of - know - so - extremely - understanding - right - confused - repetitive - kinda - not - was - like	times in a row, so it was like a little frustrating and kind of annoying.', 'It was frustrating, and then I had a breakthrough.']			
86	annoying - hate - it - irritating - kind - me - just - freaks - really - don - bothering - that - creeps - like - of - pisses - but - so - and - yos	['Um, so that has been very annoying to deal with.', 'Um, yeah, it's kind of annoying honestly.', 'And it bugs me a lot.']	122	Annoyance	feelings
87	fun - be - interesting - ll - adventure - should - gonna - will - it - an - that - think - so - hopefully - sure - re - least - looks - but - would	['Um, yeah, I expect it to be pretty fun.', 'It'll be fun.', 'Uh, I think it'll be a good chance to get out and have fun.']	115	Fun	feelings
89	prepared - feel - ready - prepare - scary - for - them - preparing - but - be - know - as - don - ll - adequately - felt - it - happen - prospects - can	['I feel prepared.', 'I prepared for it as best as I thought I could in the moment.', I've been preparing pretty well for it.']	109	Preparation	feelings

2	sleep - bed - up - woke - 00 - late - early - wake - night - go - morning - now - nap - to - at - slept - gonna - 30 - get - tired	[' I woke up at like 8:30 this morning because I went to bed at 2:00 AM last night.', "So going to bed now, g-- hopefully get a good night's sleep.", "So, like, last night I-- well, this morning I didn't go to bed until 9:00 AM, and then I woke up at 1:00."]	7898	Sleep	physical
11	sick - cold - weather - headache - throat - feeling - snow - it - the - stomach - like - still - was - really - been - my - and - is - sore - cough	["Uh, felt, felt pretty sick in the morning before it, um, though I wasn't sure if it's just part of my general being sick with a cold or something, or if it was just stress.", 'Uh, I have been feeling a little bit under the weather recently, uh, with a sore throat, so I've been trying to get as much rest as possible to, uh, accommodate for that.', "I've been pretty sick for the last couple of days."]	1626	Health	physical

15	gym - workout - exercise - the - lift - went - to - out - morning - go - good - hit - up - uh - woke - then - and - felt - exercised - exercising	['Um, I was able to go to the gym and, um-- for a couple of hours.', 'Um, at the end of the day, I went to the gym, which is cool.', 'And, um, I have to-- and I exercised a little bit today, and I will exercise tomorrow.']	1193	Exercise	physical
6	work - done - do - lot - got - get - did - stuff - much - of - didn - amount - some - have - today - as - things - more - but - to	['Got a lot of work done.', 'I was not in the mood to do much work today.', 'Um, I was-- I was able to get some done, but probably not as much as I want to do.']	2342	Schoolwork	academics
12	paper - essay - write - due - writing - on - have - my - papers - essays - finish - this - done - tomorrow - an - that - it - wrote - the - about	["And then, um, after that, I came back, and then I wrote a little of my essay 'cause I had a, a paper, not an essay, a paper due today, 15 pages.", 'The essay isn't quite due but I need to start it, so.', 'Um, my paper went good.']	1414	Paper	academics

16	productive - day - today - was - very - as - wasn - felt - pretty - been - lazy - unproductive - not - productivity - work - but - so - um - feel - like	[Um, so yeah, I also was, like, not very productive this afternoon and this evening, um, which is really stressing me out.', ' Today I felt a little bit more productive because I used a scheduling app to schedule my time.', ' I think I had a decently productive day today.']	1143	Productivity	academics
20	pset - problem - psets - math - set - cs50 - finished - cs20 - on - sets - cs51 - my - worked - finish - due - done - do - psat - the - for	['Um, I finished up my math PSet.', 'I have a pset.', 'But I finished a pset and, um, got part of another pset done.']	749	Problem set	academics
22	math - class - went - after - my - then - uh - for - to - exam - and - test - had - on - did - stuff - um - the - set - that	['It was even-- math was even worse than anthro, which I didn't think was possible.', 'Um, I like doing math.', 'I spent most of my time studying for this math mid-term that's going on.']	717	Math class	academics

25	busy - day - today - pretty - was - very - so - super - been - it - really - um - just - but - quite - of - another - uh - had - kind	['Today was a busy day.', 'Then I had a really busy morning.', ' So today was a very active day, probably one of the busiest days I had-- I've had, um, all year.']	603	Busy	academics
27	library - went - the - then - in - to - work - studied - and - some - spent - at - study - after - for - libraries - friends - worked - hours - room	['Um, in the morning, I went to the library and studied a little bit.', 'And then I went to the library.', 'Um, since then, I stayed and worked, um, in the library, which was, like, not super fun, but, like, it was good that I got work done.']	574	Library	academics
29	reading - read - readings - book - period - books - do - have - gonna - chapters - did - some - finished - done - tomorrow - the - of - week - for - to	['And, um, I actually got a lot of reading done today, which was great, and I finished, uh, b-- one-- the book I was reading, which was nice.', 'I did a little bit of reading, so I just need to do, like, all my readings this weekend.', "I	534	Reading	academics

		have a lot of reading to do, but it's good 'cause I like to read."]			
32	midterm - midterms - have - week - coming - wednesday - on - well - next - about - tomorrow - my - this - um - worried - but - that - so - nervous - had	['Um, I'm anxious about my midterms on Wednesday and Thursday.', 'I had my midterm today.', ' My main concern right now is my midterm tomorrow, and then my other two midterms on Wednesday.']	485	Midterm	academics
38	lab - partner - went - then - after - had - partners - my - labs - um - to - the - lunch - and - in - which - for - was - have - uh	['But I do have lab which is quite long.', 'And the lab didn't seem-- it wa-- it seemed more computational than wet lab, which is what I like to do.', 'Because, uh, one of my lab partners, I'm pretty sure, took pictures of our, um, results, um, in petri dishes.']	367	Laboratory	academics
44	french - linguistics - homework - language - class - my - went - essay - finish -	['I went to French class.', 'I, I work very slowly when it comes to French homework.', 'And	321	French class	academics

	for - read - did - finished - reading - then - to - have - paper - after - and	then came back and submitted my French essay.']			
52	procrastinated - procrastinating - procrastinate - procrastination - lot - work - bit - done - of - not - little - did - amount - much - but - ve - problem - kind - and - homework	['Um, but I really had already dug myself into a hole by procrastinating so aggressively, um, and just did not have enough time to do what I needed to.', 'I procrastinated a little bit more than I should have.', 'But, you know, per usual, I'm just disappointed in myself for not being able to manage my time and for always procrastinating and for never feeling motivated.']	263	Procrastination	academics
57	spanish - class - quiz - went - test - my - woke - breakfast - to - for - which - up - because - vie - est - um - exam - have - well - had	['Um, I wen-- I attempted to go to Spanish extra help today after getting my test.', 'I woke up early to study for my Spanish test.', ' So today, I had to wake up early	228	Spanish class	academics

		because I had to do a presentation for my Spanish class, which actually went really well, I think.']			
60	finals - final - my - last - are - about - have - had - home - take - first - be - coming - feel - prepared - because - for - stressed - one - over	[' So today I had my two finals.', "Um, I have two finals on Saturday, one final that I feel pretty prepared for and a final that I don't really feel that prepared for.", "So, um, wasn't really much of an eventful day, just that I had my first final, been studying for my second final that's tomorrow, and, um, yeah, I'm excited to be done with the finals tomorrow."]	223	Final exams	academics
61	expos - paper - essay - my - class - for - on - draft - went - finished - then - due - write - worked - um - and - morning - had - to - expo	['Um, in the morning, I got up and did some work for my expos class and then I went to that class.', 'And I really like this expos class, and it's making me	219	Expos class	academics

		feel more confident about my writing, something I was really scared about coming in.', 'I'm w-- I turned in my essay for my expos class on Friday, and it-- I feel like it was really bad, and I'm really worried about that class.']			
75	psych - psychology - study - abnormal - class - went - studies - after - for - course - then - section - my - had - did - homework - to - uh - the - um	['We discussed some interesting things in my psychology class.', 'Then, I went to a psych study.', 'And then I went to my psychology lecture from 1:00 to 4:00 PM which was kind of a long time.']	171	Psychology	academics
76	section - sections - went - had - then - to - after - um - was - my - which - actually - in - uh - that - gov - sectioning - and - actu - good	['So I'm hoping that, you know, with sections, one that's broken up with like 15 people-- that, you know, it's a lot more-- it's a lot easier to kind of talk and share my opinions.', 'Went to section for history, which	161	Section	academics

		was awesome, way better than I thought it was gonna be.', 'Um, at section, felt really involved in section, felt really prepared to learn and to listen, which was good.']			
13	rehearsal - choir - concert - dance - orchestra - band - had - the - went - event - then - which - and - practice - was - to - after - group - dancing - my	['After choir rehearsal, I had the auditions for solos, which I think went okay.','Um, and then I went to rehearsal, um, which was good.', 'And other than that, I had, um, dance practice, which was okay.']	1392	Rehearsal	extracurricular
26	club - extracurricular - extracurriculars - harvard - clubs - meetings - meeting - of - organization - group - for - the - this - that - in - yale - to - and - they - have	["And I did some stuff for the club that I'm in.", "Um, and I had a tryout slash interview for an extracurricular activity that I wanna do this year, um, which went really well but I still don't know if I wanna actually do it, so that's kind of up in the air.", 'I got some applications	581	Extracurricular	extracurricular

		done for some extracurriculars I want to do this semester.']			
45	interview - an - interviews - had - interviewed - job - for - article - the - interviewing - went - well - um - have - then - tomorrow - at - and - phone - that	['Um, I got up in the morning and did an interview for the newspaper over the phone for about 15 minutes.', 'And I've got an interview for a job opportunity that I really hope that I get because I have been wanting it for basically the last semester.', 'Um, my interview went fairly well.']	315	Interview	extracurricular
51	debate - tournament - team - love - round - went - for - practice - the - at - social - debating - 00 - prepping - after - to - we - like - then - prepped	['But tonight, I actually did-- we had, like, a debate with my roommates about, like, gun control in Europe versus in the US.', 'I just woke early to, um, go to the debate tournament, and, um, that was fine.', 'Um, so I went to debate.']	263	Debate	extracurricular

56	practice - training - conditioning - had - then - went - good - uh - baseball - field - workout - practiced - and - after - today - to - rehab - which - we - um	['Um, then I had practice.', 'After practice today I'm feeling physically kind of broken down a bit, but I guess that's okay 'cause coach knows what she's doing.', 'And, um, yeah, then I had practice, and practice was good.']	229	Practice	extracurricular
59	application - summer - applications - scholarships - financial - abroad - aid - program - apply - applied - for - applying - programs - study - get - thing - internships - some - the - funding	['And, um, I did some stuff for my financial aid application.', "Um, it's for the summer application that I'm trying to do for research.", "My application for the summer program I'm applying to is due tomorrow."]	225	Application	extracurricular
1	he - she - her - with - friends - him - people - like - me - friend - talked - talk - my - we - and - really - parents - them - social - to	['And, you know, just having a really good time with her and the rest of my friends there.', 'Uh, and then I called my parents on FaceTime and then called my girlfriend, and so	11841	People	social

		that was really much needed to reconnect with home.', 'And then, I hung out with my friend while she was at work and with her.']			
31	roommates - roommate - room - my - with - out - hung - our - and - we - in - um - other - of - the - friends - friend - people - one - talked	['And my roommate really wanted to hang out with people but I didn't want to.', 'Uh, so I'm glad I was able to spend some time with my roommate.', 'And, um, both of my roommates have their girlfriends over tonight, so it's kind of awkward just being in the room.']	498	Roommates	social
33	home - excited - go - be - back - socially - going - happy - to - here - now - homesick - just - looking - forward - wait - really - but - leave - um	["It's now less than a month till I get to go home, so I'm excited for that.", "Um, once again, I am home for the holidays.", "Uh-huh, I'm very happy I was able to go home for the break because it was a very fun week."]	469	Home	social

68	drunk - drink - drank - drinking - water - alcohol - drinks - didn - shelter - much - had - shift - bit - very - little - night - amount - tonight - um - homeless	["And I drank so much, like, more than I ever have before in my entire life.", "Um, well, not very drunk but I'm, I'm pretty drunk.", " I'm having a little trouble with it 'cause, currently, I'm the only sober person here."]	195	Drinking	social
70	housing - house - we - party - day - about - re - van - the - in - our - blocking - like - out - rental - of - houses - there - is - who	['Um, I'm kind of worried about housing for next year, like, not my house but, like, my roommate situation.',' Uh, not really stressed about housing day at all, which is good, which is nice.',' Today was housing day.', 'Uh, it was actually a really good day, uh, since we ended up, uh, getting our housing assignments today.']	194	Housing	social
91	party - parties - fun - went - we - was - partying - like - to - go - but - ended - the - were -	['The party was absolutely fantastic.', 'I've done a little too much partying, but had fun.',	105	Party	social

	wasn - and - then - tonight - which - really	'And the party was pretty cramped and kinda bad, honestly.']			
34	shopping - package - bought - store - went - mom - my - mall - some - and - we - clothes - black - got - then - with - to - stuff - buy - room	['Um, I went thrift shopping with my friends, which was really fun especially after, like, our heart to heart yesterday.', 'And then my mom and I went shopping.', 'And then right after that, we went to the mall, uh, where I picked up some pairs of pants, which I felt I needed.']	462	Shopping	other
39	watched - movie - watch - netflix - watching - tv - movies - show - panther - we - black - stranger - harry - potter - see - videos - saw - shows - episodes - youtube	['Um, I started watching a really good documentary series on Amazon Prime, and I've been watching that for a couple of hours today.', 'Um, oh, we also watched a movie at home, a rented one.', 'I just sat in my room and wasted time and watched TV, um, which, you know,	347	Television	other

		is always a problem for me.']			
41	survey - audio - surveys - recording - okay - yesterday - this - forgot - last - number - the - record - do - doing - did - um - night - these - so - missed	[' This is r-- this survey is really late because this is actually when I'm about to go to sleep.', 'I haven't been able to do audio surveys for the past couple days because I've been on a retreat.', ' Um, in fact, I'm supposed to be doing homework right now but I am-- instead, I am taking these surveys.']	343	Audio survey	other
42	meeting - meetings - had - then - went - 00 - to - for - the - at - and - elections - path - board - was - that - um - go - of - spot	['Then I went to a meeting and-- a couple meetings in the afternoon.', 'And then, um, I had a meeting for a little over an hour.', 'And then I went to a meeting at 5:00, which also was very good 'cause we were, like, wrapping everything up for our event that's coming up and	340	Meeting	other

		planning our next event.']			
48	shower - took - showered - take - woke - up - morning - ready - bed - and - came - then - stuff - got - go - went - back - uh - now - room	[' So today, I woke up, um, took a shower.', 'I should really, like, take a shower and go to bed, so good night.', 'Um, I came back, and I was super, super sweaty, so, and I immediately showered.']	305	Shower	other
53	blocking - block - group - mates - sushi - with - blockmates - my - we - our - is - of - who - situation - us - other - about - one - out - friends	['My roommate asked me about Housing Day and we talked about how we hope to block with each other but we don't know how it will work with our groups because we have slightly different social groups.', "So, um, just I'm stressing a lot today and, like, in the last few hours about blocking group stuff, um, and, like, who I'll be with next year,	256	Blocking group	other

		um, which doesn't-- I don't have to worry-- well, it doesn't happen formally until like February or March.", 'Um, blocking things seem to be settling out maybe, which, which is good.']			
54	birthday - friend - party - surprise - her - my - celebrate - his - birthdays - we - celebrated - was - friends - present - today - it - then - and - happy - is	[' Uh, so today was exciting because it was my birthday.', 'I also talked to my friend overseas 'cause it's her birthday, and we had a really good conversation.', 'And then at our study break, we bought a cake for my friend and sang Happy Birthday to her 'cause tomorrow's her birthday.']	248	Birthday	other
62	money - bank - financial - need - savings - finance - cryptocurrency - have - about - don - account - bill - finances - me - get -	["It's a lot of money, and not a money that I have, and so I'm worried.", "And, um, I don't know, I'm kind of, like, th-- a lot of time thinking of	217	Money	other

	summer - the - my - scholarship - is	cryptocurrency now to make sure that I don't lose any money, and, uh, yeah, that's what I'm focusing on.", "And I really don't know if I'll get this funding, and it's my last chance."]			
66	packing - pack - packed - boxes - storage - stuff - have - everything - ready - box - bag - and - all - to - need - tomorrow - water - up - home - unpacked	["Um, I've started to pack for my trip home and realized that I have a good amount of stuff.", ' So I spent most of today, um, packing.', 'I'm trying to cramp everything into one small book bag, and I definitely should have brought, like, a smaller suitcase.']	210	Packing	other
67	bus - flight - airport - ride - plane - running - on - delayed - run - we - the - got - miles - back - was - ran - took - long - then - home	['Uh, and then we got on the bus and came home.', 'Um, my flight got delayed for like 30 minutes, so, uh, I'm a little bit later than expected, but it's all good.', 'So we took a really long subway ride and	210	Transportation	other

		then, um, once we got there, we walked around.']			
69	church - went - up - go - to - morning - woke - uh - wake - ready - showered - headed - parents - got - chapel - pastor - service - head - early - picked	['Um, I woke up early because I had church this morning.', ' Okay, so I woke up this morning and, uh, I went to-- uh, uh, I got picked up by my family and went to church.', 'I went to church today, which was absolutely fantastic, and I loved it.']	195	Church	other
71	laundry - clothes - did - do - hockey - my - folded - some - game - need - wash - put - then - dry - stuff - and - dirty - to - sheets - came	['Um, I was a little irritated when I went to go do laundry 'cause a lot-- all-- basically, all the machines were taken up, but, you know, laundry has to be done.', 'Um, after that, I folded my laundry, so that was good, like, alone time for just a little bit and take a break in between my craziness.', 'I forgot to do my laundry.']	186	Laundry	other

72	brunch - went - with - friends - woke - then - up - nice - and - to - my - roommate - at - um - morning - friend - got - we - uh - some	[' Today I woke up, and I got really excited because of-- um, it's Sunday brunch.', 'Went to get brunch with my girlfriend and her cousin.', 'Uh, ended up getting some brunch alone.']	181	Brunch	other
79	break - breaks - take - gonna - home - tomorrow - this - going - be - need - over - like - go - just - forward - needed - have - myself - looking - to	["Um, and, yeah, it was still nice to sort of finish up something and then take a-- take a well-earned break.", "It's like you never have time to yourself, and every time you try to take a break, you just feel guilty for it.", 'But I honestly could've used a break today, and I did not get a break.']	155	Break	other
82	decision - choice - successful - decisions - one - no - best - made - decided - have - either - it - success - you - tough - that - think -	['Um, and I think I made the right decision.', 'So that was, like, not really the best decision, I think.', 'It was a good decision I think.']	129	Decision	other

	don - haven - stop				
83	cleaned - clean - room - cleaning - desk - my - stuff - up - vacuumed - clothes - everything - mess - area - packing - and - better - is - um - decluttered - vacuum	['Um, then I vacuumed, made me feel a lot better.', 'Um, I'm cleaning my sheets, cleaning my room, Lysol, and everything, um, so hopefully, that kind of helps.', 'Um, I haven't cleaned my room recently, which is kinda bugging me, but, you know, I'll get it done soon.']	128	Cleaning	other
90	hair - haircut - washed - cut - my - shampoo - wash - got - braided - eyebrows - long - shaved - did - put - ear - and - hairspray - moisturized - braids - haircuts	["And then I got, got my hair cut finally, which was nice since I haven't done that since like before winter.", 'And then I went and got my haircut because every time that my mom sees me, she reminds me that I haven't gotten a haircut in over a year.', 'Did my hair, a little bit of self-care.']	108	Grooming	other

5	yeah - don - um - know - but - let - so - that - about - say - much - why - it - what - yep - see - oops - knows - else - uh	['So yeah, um, that's pretty much it.', 'Uh, yeah, that's what's going on right now.', 'So yeah.']	2585	Yeah	filler
8	redacted - to - at - went - the - in - from - for - then - we - and - friend - um - with - which - back - my - of - was - uh	['Tomorrow, I'm going to go into [redacted] with my roommate and, um, go to a concert, which should be really fun.', "Um, and it kind of distracted me from things that have been going on at [redacted].", 'Um, tomorrow we have another full day of competition, and then we're flying back to redacted from the [redacted] Airport, but I feel like it's been a really good time too getting to know the members of the team and to see a really cool city here in [redacted]."']	1931	Redacted	filler

10	see - ll - let - goes - hopefully - forward - be - looking - we - how - will - happens - fine - it - okay - what - but - well - hope - gonna	["I'll see what happens.", "We'll see what happens about that.", "Um, so we'll see how that goes."]	1631	We'll see	filler
21	bye - night - goodnight - good - okay - signing - off - goodbye - anyway - thank - laughter - yeah - merry - peace - yay - right - you - sign - christmas - so	['Okay, bye.', 'Okay, bye-bye.', 'Good-bye.']	719	Bye	filler
23	don - know - knowest - well - not - we - but	["I don't know.", "Don't know.", "I do not know."]	669	I don't know	filler
24	do - what - did - don - know - that - can - gonna - fix - it - problem - wanna - how - ll - best - well - doing - just - so - need	["Um, and I just don't know what to do.", 'Um, and then I-- what did I do?', 'Uh, so that's what I did.']	606	Did that	filler
37	hello - hi - fun - don - know - good - was - so - it - that - dunno - okay - suite -	['Hello.', 'Okay, hello.', ' Hi.']	391	Hello	filler

	enjoyable - guys - hmm - anyway - again - home - from				
40	like - okay - oh - said - you - was - what - no - wow - comfort - zone - yes - yeah - just - and - whatever - something - this - hmm - god	['And they said, "Oh, okay.', 'And I was like, "Oh, okay.', 'And I was like, "Okay, you know what?']	344	And I was like	filler
50	else - what - yep - um - yes - okay - hmm - anything - anyway - nothing - sure - and - next - oh - who - last - two - see - so - things	['Um, what else?', 'What else?', 'Okay, last two things.']	285	What else	filler
63	laughter - laugh - laughed - yeah - don - laughing - like - know - it - but - you - so - good - just - that - about - oh - of - guess - um	['[laughter] That was actually kind of funny.', 'Um, I couldn't stop laughing about it originally because it's so ridiculous, but I don't know what to do about that.', 'I'm like having a fun time laughing with my friends about it and so it also like made me laugh a	215	Laughter	filler

		lot, so that's good.']			
65	okay - hold - fine - please - well - held - on - no	['Okay.', 'Well, okay.', 'Hold on.']	212	Okay	filler
78	right - all - short - keep - cool - late - super - doing - one - gonna - it - this - but - so - - - - - -	['All right.', ' All right, cool.', ' So, all right.']	157	Allright	filler
80	happened - let - that - important - tch - yeah - happen - think - what - so - me - um - stupid - didn - sounds - but - all - colon - remember - it	['Um, so that happened.', 'Not really sure how it happened, but, um, yeah.', 'Um, yeah, so that, that happened.']	146	That happened	filler
84	whatever - don - know - and - dunno - anyhow - whatnot - so - yeah - just - - - - - - - - - -	['Whatever.', 'Just, whatever.', 'So, whatever.']	127	Whatever	filler
88	done - get - almost - ll - it - hopefully - but - gonna - be - mean - now - that - sure - all - with - somehow -	['But I'm gonna say-- uh, I'm just gonna have to get that done, uh, sooner or later.', "Um, but yeah, I mean, what's	110	Done	filler

	least - need - able - so	done, it's done.", 'I'm so done.']			
-1	and - to - that - it - the - my - um - like - of - so - just - for - then - was - but - uh - really - in - went - on	['Um, I got up and went to breakfast and then did some work for my classes.', "And so I just-- I just don't know what to do.", 'I had a nice time, but then I was kinda stressed because I went back to my room and we hung out a while when I actually wanted to do some readings for some classes.']	47799	DNU	DNU