Using natural language processing to understand changes in emotions and psychopathology: Evidence from observational cohorts and a clinical trial

Lorenzo Lorenzo-Luaces, PhD[1]

Lauren A. Rutter, PhD[1]

Meghana Boinpally, MS[2]

Jacqueline Howard, BA[3]

Danny Valdez, PhD[4]

Johan Bollen, PhD[2]

1 – Psychological and Brain Sciences, Indiana University-Bloomington, Bloomington, IN, 47401

2 – Luddy School of Informatics, Computer Science, and Engineering, Indiana University-Bloomington, Bloomington, IN, 47401

3 – TRAILS to Wellness, Ann Harbor, Michigan, 48105

4 – School of Public Health, Indiana University-Bloomington, Bloomington, IN, 47401

Abstract

Introduction: Natural language processing (NLP) metrics have emerged as a way of understanding emotions. We explored the application of NLP to understand emotion and emotion regulation in psychopathology in two studies In Study 1, we analyze social media messages from the Studies of Online Cohorts for Internalizing symptoms in Language (SOCIAL), two observational cohorts of social media users (*N*=3,334). In Study 2, we analyze data from a cognitive restructuring task embedded in the cognitive-behavioral therapy condition (*n*=409) of a large treatment trial (*N* = 828). **Methods**: In Study 1, we obtained Valence Aware Dictionary for sEntiment Reasoning (VADER) scores for 1,471,079 million Tweets. We used median VADER scores and variability in VADER scores to predict different dimensions of psychopathology. In Study 2, we use VADER and Sentiment Analysis and social Cognition Engine (SEANCE) to analyze 28,986 words from the cognitive restructuring task to assess whether NLP metrics are indicators of emotion regulation in the task, are correlated to individual differences, and can predict positive (well-being) and negative (depression) mental health at weeks 2 and 8. **Results:** In Study 1,affect variability, especially negative affect was associated with more severe Internalizing-fear and distress and Somatoform-distress symptoms and less severe Externalizing-substance use symptoms. In Study 2, NLP metrics improved during the cognitive restructuring task, were associated with self-report use of cognitive reappraisal, emotional stability, and extraversion, and predicted symptom change post-intervention. **Discussion:** These studies illustrate the potential to use NLP for understanding changes in emotional dynamics, including minute-scale changes in emotions during psychological interventions.

Keywords: Transdiagnostic, machine learning, natural language processing

Public Health Significance

Text collected from social media or digital mental health interventions was associated with individual differences in mental disorder symptoms, including the different constellations of symptoms endorsed (e.g., internalizing vs. externalizing). The text also tracks improvements during a mental health intervention and after the intervention. These findings provide novel ways to characterize psychopathology from a transdiagnostic perspective.

Using natural language processing to understand changes in emotions: Evidence from observational cohorts and a clinical trial

Mental disorders, including depression, anxiety, substance use, and pain-related conditions, account for a substantial proportion of disability attributed to illness worldwide (Murray et al., 2012). According to hierarchical models of psychopathology, like the Hierarchical Taxonomy of Psychopathology (HiTOP, Kotov et al., 2017), most of these clinical problems can be grouped into three superspectra that capture symptoms of (1) emotional dysfunction (Watson et al., 2022), including internalizing symptoms and somatoform conditions, (2) externalizing symptoms (Krueger et al., 2021), including substance use and antagonism, and (3) psychosis (Kotov et al., 2020), including thought disorder symptoms and detachment. While the HiTOP model groups these superspectra of symptoms separately, they are also known to interact (e.g., substance use increases risk of internalizing symptoms).

Psychopathology can be understood as the product of difficulty using adaptive emotion regulation strategies and the overuse of maladaptive coping strategies (Jazaieri et al., 2013). These failures to regulate emotions may be manifested in affect variability, or the degree to which affect changes over time (Kuppens et al., 2007). Meta-analyses of emotion dynamics suggest that affect variability may be associated with worse mental health in general (Houben et al., 2015). While fluctuations in emotions are typical in daily functioning, more variation has been linked to more severe symptoms of anxiety and depression (Gruber et al., 2013), poorer sleep (Leger et al., 2019) and even lowered immune functioning (Jenkins et al., 2018). Positive and negative affect variability may differentially predict psychopathology concurrently and prospectively (Scott et al., 2020; Sperry et al., 2020). For example, externalizing disorders may be associated with lower positive affect inertia and more positive affect variability, while internalizing disorders may be associated with higher negative affect variability and lower positive affect variability (Scott et al., 2020).

This conceptualization of psychopathology as failure to regulate emotions is supported by clinical research. Meta-analyses of emotion regulation experiments (Webb et al., 2012) suggest that regulating emotions by cognitive reappraisal, cognitively reaxamining emotional stimuli to change their meaning, is associated with improvements in emotions (standardized mean difference (SMD) = 0.45, 95% CI: 0.35, 0.56). By contrast, expressing emotions (i.e., not suppressing their expression) is also associated with emotion outcomes (SMD = 0.10, 95% CI: 0.01, 0.18). Psychological interventions, the most widely studied of which are cognitive-behavioral therapies (CBTs), may improve psychopathology partly by improving the capacity to regulate negative emotions through strategies like cognitive reappraisal (Lorenzo-Luaces et al., 2015, 2016) and decreasing maladaptive strategies like suppressing the expression of emotions (Chawla & Ostafin, 2007).

While difficulty regulating emotions may be a shared feature of different constellation of mental health symptoms, symptoms of psychopathology are very heterogeneous in their presenting features (Kotov et al., 2017; Lorenzo-Luaces, Buss, et al., 2021) as well as their longitudinal courses (Lorenzo-Luaces, 2015). Research in clinical psychology typically assesses emotions via self-report measures. Self-report has many appealing qualities for the study of emotions. For example, if one assumes that individuals have access to their emotional states, then self-report assessments should have high levels of face validity. Additionally, self-reported emotion assessments have been shown to have predictive validity (e.g., Lorenzo-Luaces, Peipert, et al., 2021). Despite the promise of self-report, it has a variety of limitations. For example, self-report imposes a burden on participants. Assessing multiple constructs is impossible without burdening participants (i.e., including more questions). When research on emotions and emotion dynamics is conducted longitudinally, there are concerns about the extent to which repeated administration of questionnaires maintains a measurement of the same underlying constructs (i.e., measurement invariance, Fried et al., 2016).

Natural language processing

Natural language processing (NLP) is a subfield of computer science concerned with extraction of data from text. One common use of NLP in clinical psychology is to capture the polarity of a text (i.e., whether it conveys negative vs. neutral vs. positive emotions). Perhaps the most widely used NLP tool in clinical psychology is the Linguistic Inquiry and Word Count (LIWC, pronounced “Luke,” Kahn et al., 2007). For a piece of text (e.g., a text entry on an automatic thought record), LWIC analyzes the percentage of words that belong to specific categories. LWIC outputs metrics of linguistic (e.g., first-person pronouns), psychological (e.g., negative emotion words), and thematic categories (e.g., leisure, money) metrics. For example, for the 3-word sentence “I hate myself,” LWIC may give a score of 0.66 for “first person pronouns” (i.e., “I” and “myself”) a score of 0.33 for “negative emotion words” (i.e., “hate”), and a score of 0 for “positive emotion words.” For the 3-word sentence “I love myself,” LWIC may give a score of 0.66 for “first person pronouns” (i.e., “I” and “myself”) a score of 0 for “negative emotions”, and a score of 0.33 for “positive emotions” (i.e., “love”). The 8-word sentence “I dislike myself sometimes, but I am good” may receive a score of 0.38 for first-person pronouns (i.e., “I”, “myself”, “I”), a 0.12 for negative emotions (i.e., “dislike”), and a 0.12 for positive emotions (i.e., “good”).

LWIC scores are often used to make assumptions about the psychological processes that may have produced the texts. For example, using the percentage of negative emotion words as an index that captures feeling negative emotions, “I hate myself” is more negative than “I dislike myself but I am trying,” which is itself more negative than “I love myself.” LWIC scores have been widely used in psychology and related fields. In 2010, Tausczik et al. (2010) identified over 100 peer-reviewed articles that have correlated LWIC-derived metrics and correlated them to psychological constructs. For example, greater use of personal pronouns (e.g., “I”) in text has been correlated a greater severity of depression symptoms (see Holtzman et al., 2017 for a review), a finding that connects with research using cognitive tasks linking depression to increased self-referential processing (LeMoult et al., 2017). Similarly, greater use of negative emotional words, including those expressing depressive symptoms, appears related to depressive symptoms (JH Balsters et al., 2012).

One appealing feature of LWIC is that it is relatively easy to understand and calculate. Nonetheless, LWIC has some notable limitations. One limitation is LIWC uses a predefined, and proprietary,lexicon, or dictionary of words, that map on to the different LWIC categories. Although the LWIC dictionary contains thousands of words, it probably does not capture the full range of linguistic and emotional expressions that are possible to extract from a text. Another limitation is that LIWC is based on a rather surface-level feature of language: word frequency. Thus, traditional LWIC output does not account for the context in which words are used, and thus ignores important components of how meaning and feelings are expressed in language. For example, although LWIC counts negations (e.g., the number of “not” words) LWIC does not integrate that information in its evaluation of LWIC categories. In the text “I am not happy,” the word “happy” would be counted as a positive emotion word (i.e., 0.25 positive emotion) even though the intent of the word in context is to indicate negative sentiment. Another limitation of

LWIC relevant to the study of emotions is that LWIC was not developed specifically to quantify the emotional state expressed in a text or its overall sentiment.

Other more complex NLP algorithms have been specifically designed for the purposes of understanding the sentiment in text. For example, Valence Aware Dictionary and sEntiment Reasoner (VADER) is an NLP tool designed to quantify the overall sentiment expressed in a text (Hutto & Gilbert, 2014). In addition to providing information about polarity (e.g., negative vs. positive), VADER also contains information about the intensity of the sentiment which it measures on a scale of -1 (extremely negative) to +1 (extremely). For LWIC, “I hate him” and “I dislike him” may have a similar negative emotion score: 0.33 due to the one word indicating dislike/hate. VADER was developed using a “wisdom of the crowds” approach to pool ratings of the different levels of emotionality (sentiment) associated with different terms. VADER also attempts to account for the context in which words are used by altering the sentiment expressed based on punctuation (e.g., “I hate him!!!” has a VADER score of -0.68), capitalization (“I HATE him!!!” has a VADER score of -0.74), degree modifiers (e.g., “I absolutely HATE him!!!” has a score of -0.76), shifts in polarity due to contrastive conjuction (e.g., the use of “but” as in “I REALLY HATE HIM but he has a point” produces a VADER score of -0.58), and shifts in polarity due to negation. In its development paper, VADER was compared to LWIC and 10 other NLP metrics and found to better approximate the sentiment of text as rated by humans.

These and other advanced NLP methods have been used to study vulnerability to psychopathology (e.g., Bollen et al., 2011). For example, Bathina et al. (2021) measured lexical proxies of cognitive distortions (e.g., “should,” “must,” “have to,” “nobody,” “always”) a concept from CBT which points to rigid or inflexible thinking (Lorenzo-Luaces et al., 2015, 2016). As suggested by the Cognitive Model underlying CBT (Beck & Haigh, 2014), individuals with depression made more use of cognitive distortions than a random sample of individuals (Bathina et al., 2021; see also Al-Mosaiwi & Johnstone, 2018). Thus, NLP tools have demonstrated applicability to the study of emotions as well as to constructs relevant to emotion processes like the cognitive rigidity associated with psychopathology.

Current study

While previous studies have correlated NLP metrics to self-reported assessments of psychopathology (e.g., depression), a notable limitations of many of these studies, including our own prior work, is that they employed a DSM-based classification of psychopathology, usually relying on one constellation of clinical symptoms (e.g., depression). More modern, data-informed models like the HiTOP highlight that the co-variation between seemingly diverse symptoms of psychopathology may help us understand shared psychological processes (Kotov et al., 2020; Krueger et al., 2021; Watson et al., 2022). Moreover, many of these studies have relatively small samples (e.g., Liu et al., 2022). In the current study, we explore the use of NLP to understand emotion in the context of psychopathology deriving text data from two unique contexts: social media (Study 1) and a digital mental health intervention (Study 2).

Study 1: Studies of Online Cohorts for Internalizing symptoms and Language (SOCIAL)

Social media is a hallmark of modern life and thus an important context to understand. In the United States, for example, over 70% (Perrin, 2015) of individuals belong to at least one social media platform. Twitter is used by about 23% of U.S. adults and half use the platform at least daily. The SOCIAL studies are a series of cohorts in which we triangulate self-reported questionnaires probing symptoms of internalizing, externalizing, and somatoform diagnoses (see Table 1) with data derived from the social media website Twitter. Both studies were approved by the Indiana University IRB (2002549202, 2005948214). The SOCIAL studies were originally described elsewhere (CONCEALED FOR PEER REVIEW; CONCEALED FOR PEER REVIEW). In the initial report, we described that the differences between individuals who donated their social media accounts and those who did not were few, were usually small in magnitude, and usually related to atypical response patterns in individuals who were identified as having provided “bot-like” social media accounts (CONCEALED FOR PEER REVIEW).

Participant recruitment

SOCIAL-I purposely sampled Twitter users via Qualtrics panels. We aimed to recruit around 1,000 Twitter users. The sample was selected to represent the U.S. on the intersections of age, gender, and race/ethnicity. SOCIAL-II recruited college students from a predominantly White university in the Midwest. Individuals were recruited from September 2020 to the present with no end recruitment goal and were compensated for credit in an introductory psychology course.

Measures

For individuals in SOCIAL-I and SOCIAL-II we collected information on:

Demographics. We collected age, race/ethnicity, sex assigned at birth (male, female, other/inconclusive, or prefer not to say), gender identity (male, female, non-binary, genderqueer, or agender, other, or prefer not to say), and sexual orientation (heterosexual/straight, homosexual/gay, bisexual/pansexual, other, prefer not to say).

Mental health. We compiled a battery of self-report disorder screening questionnaires of psychopathology (see Table 1). These measures were chosen because A) they measure symptoms that are relatively common (e.g., depression) or relatively uncommon but highly impairing (e.g., drug use), B) are indicators of some of the major domains of psychopathology as per contemporary nosologies (e.g., Kotov et al., 2017), C) were freely available, and D) are widely used. Most of the measures we used are the *DSM* severity measures recommended by APA (e.g., social anxiety, panic, worry, substance use) or were measures that eventually were adapted into *DSM* severity measures (e.g., the PHQ-9 and PHQ-15 for depression and somatic symptoms, respectively).

Table 1. Assessment of psychopathology for Study of Online Cohorts for Internalizing symptoms and Language (SOCIAL−I =1123; SOCIAL−II = 1988)

Construct	Measure	Items	Response options	Original range	Omega
Internalizing					
Depression	PHQ-9	9	0 - 3 (not at all - nearly every day)	0 - 27	0.89
Social anxiety	DSM Severity	10	0 - 4 (never -all of the time)	0 - 40	0.94
Panic	DSM Severity	10	0 - 4 (never -all of the time)	0 - 40	0.95
Worry	DSM Severity	10	0 - 4 (never -all of the time)	0 - 40	0.93
Somatoform					
Pain	PHQ-15	15	0 - 2 (not bothered a lot - bothered a lot)	0 - 30	0.85
Insomnia	ISI	7	Varies per question	0 - 28	0.87
Externalizing					
Alcohol use	AUDIT	10	Varies per question	0 - 40	0.90
Substance use	DSM Severity	10	0 - 4 (not at all - nearly every day)	0 - 40	0.82

Note: PHQ = Patient Health Questionnaire, DSM Severity = DSM Severity Measure for each diagnosis, ISI = Insomnia Severity Index, AUDIT = Alcohol Use Disorder Identification Test

Social media. We asked for a Twitter handle for all individuals in the study who identified that they were Twitter users. The Twitter Application Programming Interface (API), a free and public interface provided by Twitter, provides access to a record of an individual’s past tweets (“timelines”).

Botometer

We assessed whether the corresponding Twitter accounts were valid and belonged to real users using the Botometer API (Yang et al., 2022). BotOMeter, formerly “Bot or Not” is a social media analysis tool that uses machine learning algorithms to assess the likelihood that a given Twitter account is a bot, as opposed to being operated by a human. It takes various factors into consideration, such as the account’s activity patterns, language use, and relationships with other accounts, to generate a score that indicates how likely it is that the account is a bot. As per the recommendations of the Botometer developers, we explored the distribution of bot scores and used a cut-off of 0.42 to classify accounts as “bot-like.” Bot-like accounts were not used in our analyses of social media texts.

VADER

We calculated VADER scores for each “tweet” obtained from valid users. The compound VADER scores range from -1 to +1 with a greater score indicating a more positive sentiment and the sign of the score indicating polarity (e.g., negative means a negative sentiment). VADER also provides specific scores of negative and positive sentiment. For each user we obtained median VADER scores and the standard deviation of the scores. We employ the standard deviation as an easily-communicable measure of emotion variability.

Analytic plan

All analyses were conducted using the R programming language (R Core Team, 2021, version 4.1.2) in R Studio (RStudio Team, 2020). For continuous variables, we provide descriptive statistics in the form of means (M) and standard deviations (SDs). For categorical variables, we present frequencies and percentages.

Given the wealth of measures we employed (see Table 1), we used an exploratory factor analysis (EFA) with the *R* package *psych* to identify common dimensions across the measures. To decide the number of factors to extract, we first used a parallel analysis in *psych* with 5,000 simulated samples. We used the number of factors for the parallel analysis to extract factor scores from an EFA of all the self-reported measures. For the purposes of having more power, we conduct these analyses with the whole sample (*N* = 3334).

Next, we focus on the subsample of SOCIAL participants who a) were Twitter users, b) who provided their Twitter handle, and c) who had low BotOMeter scores. As per our previous study (CONCEALED FOR PEER REVIEW), these participants are not substantively different than individuals who denied being Twitter users or who did not provide their handles. We focus this analysis on individuals from whom we could retrieve at least 30 tweets. Finally, to assess how dimensions of psychopathology are related to emotions and emotion variability online, we conducted two analyses. First, we regressed the median VADER score on the dimensions of psychopathology extracted from the EFA. Because the median scores were distributed like a count, we used a logistic regression for these purposes. To assess the relation between affect variability and dimensions of psychopathology, we regressed the standard deviation of VADER scores per user on their factors scores, extracted from the EFA. To minimize the potential influence of outliers (e.g., individuals with extremely variable affect), we used a robust linear regression approach with the *R* package *MASS*. *gtsummary* and *flextable* were used to create all tables in Study 1 and Study 2.

Transparency and openness

All data and analysis code are made available in our Open Science Framework (https://osf.io/hvut4/). We report for Study 1 and Study 2 (see below), entry criteria, relevant study design features (e.g., sample size planning), and the R packages required to replicate the analysis.

Table 2. Assessment of psychopathology for Study of Online Cohorts for Internalizing symptoms and Language

Characteristic	Overall, N = 3,334[1]	SOCIAL-I, N = 1,123[1]	SOCIAL-II, N = 2,211[1]
Age	24.47 (10.82)	34.70 (12.81)	19.07 (2.82)
Missing	225	50	175
Race/ethnicity			
Asian	299 (9.2%)	62 (5.5%)	237 (11%)
Hispanic	201 (6.2%)	87 (7.8%)	114 (5.3%)
MENA	2 (<0.1%)	0 (0%)	2 (<0.1%)
Non-Hispanic Black	288 (8.8%)	128 (11%)	160 (7.5%)
Non-Hispanic White	2,384 (73%)	826 (74%)	1,558 (73%)
Other	82 (2.5%)	19 (1.7%)	63 (3.0%)
Missing	78	1	77
Sex assigned at birth			
Male	951 (29%)	485 (43%)	466 (21%)
Female	2,303 (69%)	636 (57%)	1,667 (75%)
Genderqueer/Refused	80 (2.4%)	2 (0.2%)	78 (3.5%)
Twitter user	2,454 (74%)	1,123 (100%)	1,331 (61%)
Missing	29	0	29
Depression (PHQ-9: 0-27)	7.85 (6.26)	9.01 (6.93)	7.25 (5.79)
Missing	67	3	64
Social anxiety (DSM: 0-40)	9.16 (9.14)	11.14 (10.42)	8.13 (8.21)
Missing	70	3	67
Generalized anxiety (DSM: 0-40)	9.74 (8.78)	11.50 (10.38)	8.81 (7.65)
Missing	76	4	72
Panic (DSM: 0-40)	5.48 (8.06)	8.83 (10.17)	3.73 (6.00)
Missing	72	4	68
Pain (PHQ-15: 0-30)	7.75 (5.61)	9.12 (6.64)	7.04 (4.84)
Missing	89	14	75
Insomnia (ISI: 0-28)	9.96 (5.94)	11.60 (6.45)	9.10 (5.46)
Missing	63	1	62
Alcohol use (AUDIT: 0-40)	5.92 (6.95)	8.11 (9.19)	4.77 (5.05)
Missing	77	0	77
Drug use (DSM: 0-40)	1.82 (4.55)	3.85 (7.06)	0.77 (1.52)
Missing	70	6	64

[1] Mean (SD); n (%)

Note: PHQ = Patient Health Questionnaire, DSM = DSM Severity Measure for each symptom, ISI = Insomnia Severity Index, AUDIT = Alcohol Use Disorder Identification Test

Results

Demographics for participants are reported in Table 2.

Psychopathology indices in SOCIAL Studies

The parallel analysis of the 8 measures suggested the presence of 3 factors. We dubbed the first factor *Internalizing-fear and distress* as it had moderate to strong (0.55-0.83) loading from all measures of internalizing symptoms (i.e., PHQ-9 and the DSM Severity Measures for Social Anxiety, Panic, and Generalized Anxiety), along with a moderate loading (0.55) of pain on the PHQ-15. The second factor had strong loadings from alcohol use on the AUDIT (0.63) and drug use on the DSM Severity Measure (0.85) and was dubbed *Externalizing-substance use*. The third factor we labelled *Somatoform-distress* as it had strong loadings from insomnia on the ISI (0.78), pain on the PHQ-15 (0.55), and depression on the PHQ-9 (0.65).

Tweet extraction and BotOMeter analysis

Of the participants in SOCIAL-I and SOCIAL-II, all in SOCIAL-I (*n*=1,123) and about two thirds in SOCIAL-II reported being Twitter users (n = 1,331). 1530 individuals (SOCIAL-I: n=776, SOCIAL-II: n=754) provided their Twitter handles (65.21%). A fraction of the accounts had high BotOMeter Bot Scores and were excluded from (SOCIAL-I: 22.29% (n=173), SOCIAL-II: 30.24% (n=228)). We also removed 25 individuals who Tweeted less than 30 times leaving us with a sample of 1076 individuals. On average, we extracted 1367.17 tweets per person who volunteered their handles (SD = 1180.24, Me = 910, IQR = 276.75, 2661.5). Across these samples, we extracted 1,471,079 million Tweets.

Affect and affect variability

The median VADER score per user indicated a neutral sentiment (M = 0.08, Median = 0) although the standard deviation per user suggested that there was variability in the VADER scores (M = 0.41, SD = 0.05) as well as in their negative (M = 0.11, SD = 0.03) and positive constituent scores (M = 0.17, SD = 0.04). The median VADER score appeared unrelated to scores on the Internalizing-fear and distress (OR = 0.91, 95% CI: 0.78, 1.05), Somatoform-distress (OR = 0.99, 95% CI: 0.85, 1.16), or Externalizing-substance use (OR = 1.08, 95% CI: 0.93, 1.23) factors. However, affect variability was lower for individuals scoring higher on the Externalizing-substance use factor (β = -0.08, 95% CI: -0.13, -0.03) and higher for individuals scoring higher in the Somatoform-distress (β = 0.07, 95% CI: 0.02, 0.12) and Internalizing-fear and distress factors (β = 0.05, 95% CI: 0.00, 0.10) (see Figure 1)

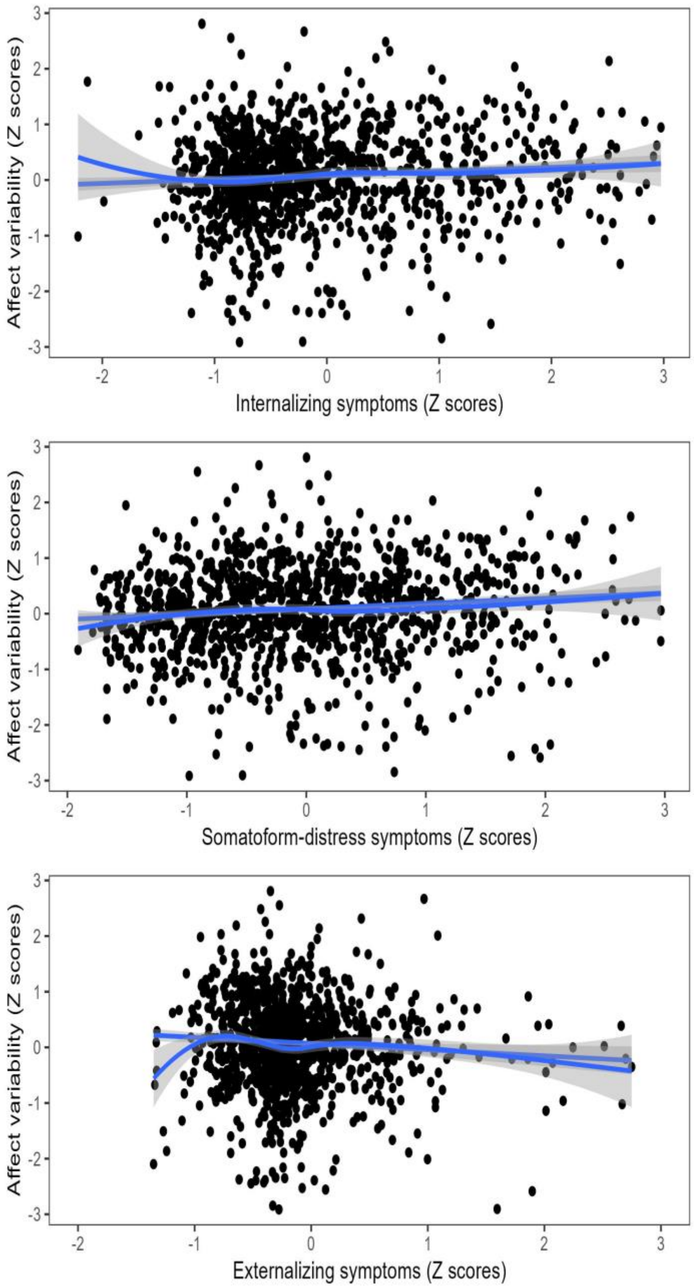

Figure 1. Association between dimension of psychopathology and affect variability across 1,471,079 million Tweets for 1,076 individuals

Because we analyzed data from two cohorts that were relatively different in how they were sampled, we explored whether there were statistically-significant interactions between the study cohort (i.e., SOCIAL-I vs. SOCIAL-II) and the factors of psychopathology in predicting VADER scores or their variability. There were no statistically-significant interactions between the study cohort and the symptom dimensions in predicting VADER scores or their variability (all *p*s > 0.14). Finally, because the VADER scores are computed by weighting negative language vs. more positive language, we examined whether it was positive sentiment or negative sentiment variability that was associated to Internalizing-fear and distress, Externalizing-substance use, and Somatoform-distress dimension of psychopathology. Interestingly, even though there was somewhat more variability in the positive VADER sentiment than the negative sentiment, it was negative sentiment variability that was more strongly associated with the dimensions of psychopathology (see Table 3). Positive sentiment variability was also lower for individuals higher in Externalizing-substance use.

Table 3. Associations between variability in negative vs. positive sentiment and dimensions of psychopathology across 1,471,079 million Tweets from 1,076 individuals

	Negative affect		Positive affect	
Characteristic	Beta	95% CI[1]	Beta	95% CI[1]
Internalizing-fear and distress (Z scores)	0.08	0.02, 0.14	-0.02	-0.08, 0.03
Externalizing-substance use (Z scores)	-0.11	-0.17, -0.05	-0.07	-0.13, -0.02
Somatoform-distress (Z scores)	0.07	0.01, 0.13	-0.03	-0.08, 0.03

[1] CI = Confidence Interval

Discussion

We surveyed social media users with a battery of measures of psychopathology and accessed and analyzed the content posted from their social media accounts. Our work extends previous efforts with a more modern conceptualization of psychopathology: the HiTOP model. Our results are broadly consistent with the latest iteration of the HiTOP model which subsumes internalizing and somatoform symptoms under an “emotional dysfunction” super-spectrum. Not only were internalizing symptoms of fear and distress strongly correlated with pain but symptoms of pain and insomnia very strongly correlated with depression. Moreover, both sets of symptoms showed the same association with negative affect variability: greater variability was associated with more severe symptoms.

Some limitations of our data are worth considering. First, individuals post content on social media for different motives (Luchman et al., 2014; e.g., Riley et al., 2022) leaving open the possibility of various biases in the collection of data. Just because an individual does not post negative content it does not imply that they are not experiencing negative emotions. While BotOMeter and VADER are powerful tools to study online behavior, it is plausible to think that different results could be obtained with different NLP algorithms. Finally, while our data speak to the ability of NLP metrics to capture individual differences in psychopathology, it is unclear how meaningful NLP metrics are above and beyond capturing individual differences in symptom severity and the constellation of symptoms endorsed. In Study 2, we expand on these limitations by focusing on text responses to a uniform text prompt in a treatment context. Additionally, we introduce other NLP metrics that go beyond sentiment and emotions to capture cognitive and behavioral constructs related to CBT.

Study 2: Applicability of NLP in a treatment study

We analyze data from a recently completed randomized controlled trial comparing the online single session intervention the Common Elements Toolbox (COMET, CONCEALED FOR PEER REVIEW) to a waiting list control. COMET is a self-guided 4-module intervention designed based on two core principles of CBT: cognitive restructuring and behavioral activation, and two of positive psychology: gratitude and self-compassion. Prior work supporting the efficacy of COMET include an open trial (Wasil et al., 2021) and an RCT (Wasil et al., 2022). COMET presents individuals with psychoeducation in the form of texts and brief exercises. Here, we focus on the cognitive restructuring task embedded in COMET given the consistent association in experimental studies of cognitive reappraisal and emotion regulation (Webb et al., 2012). Specifically, our aims with Study 2 were to evaluate NLP metrics as: indicators of the success regulating emotions following the cognitive restructuring (Aim A), correlates of individual differences (Aim B), and predictors of depression and well-being following COMET at a 2-week and 8-week follow-up (Aim C).

Methods

This is a randomized controlled trial (RCT) approved by the institution’s internal review board. There was a pre-registration for the design of the study (CONCEALED FOR PEER REVIEW). The trial was powered to detect rather small differences (SMD=0.20) between

COMET and control. In the current analyses we only use the COMET arm of the RCT (N=409) because individuals in the waiting list control did not provide written text responses.

Participants

Online workers living in the United States were sampled from Prolific.co, an online platform that hosts research studies. The only inclusion criteria we used for the trial was an affirmative answer (i.e., responding "yes") to the question "Do you have – or have you had – a diagnosed, on-going mental health/illness/condition?" No explicit exclusion criteria were used. However, it could be argued that relatively steady internet access, and internet literacy, was an entry criterion because the study required individuals to log in to Prolific 1-4 times. Of the 409 participants, 387 (94.62%) of individuals provided text responses. We focus on these individuals.

Cognitive restructuring task

The cognitive restructuring task embedded in COMET is a simplified version of an automatic thought record. Individuals were given these instructions:

"The ABCD technique is most helpful when we find ourselves stressed or upset, and we suspect that we might not be thinking clearly or objectively about ourselves or the situation we find ourselves in. It involves the following steps: A) Activating Event: Identify the objective facts about a situation that you find stressful or upsetting, B) Beliefs: Identify the thoughts and beliefs that are going through your head when you think about the stressful situation, C) Consider: Consider your situation from different points of view, and D) Debrief: Reflect on how these new explanations/points of view impact your mood and feelings.

After this introduction, individuals read an example of an individual using the ABCD technique. After this, individuals are invited to identify a situation that is upsetting to them and apply the ABCD technique. We extracted the text from part B of the task, where the individual identifies their automatic thoughts about the task, and part C when the individual has completed the cognitive restructuring.

Before and immediately after the task, participants completed a visual analogue scale (VAS: 0-100) measuring distress in response to the prompt: “How are you feeling about your situation on a scale of 0-100 (0 = not upset at all and 100 = very upset)?”

Measures

Depression was measured with the PHQ-8, which is the PHQ-9 (see Study 1 Methods) minus the question about death ideation and self-injury. However, we prorated scores on the PHQ-8 so the range (0-24) would be in the PHQ-9 metric (0-27). Other measures administered are described below.

Measures were administered at baseline (PHQ-9, GAD-7, WHO-5, ERQ, TIPI), and two, four, and eight weeks post-baseline (PHQ-9, GAD-7, WHO-5, ERQ).

Generalized Anxiety Disorder Scale 7 (GAD-7). The GAD-7 (Spitzer et al., 2006) is a 7-item self-report questionnaire assessing the frequency of anxiety symptoms. It was developed as a screening tool for primary care. Responses range from 0 (“Not at all”) to 3 (“Nearly every day”), producing scores from 0-21 with greater scores indicating more frequent depressive symptoms. At baseline, the GAD-7 appeared to be an internally consistent measure of generalized anxiety (ω = 0.91, 95% CI: 0.90, 0.92).

WHO Well-being Index (WHO-5). The WHO-5 (Topp et al., 2015) is a five-item self-report scale that measures subjective well-being, an aspect of positive mental health. Its items are rated on a scale of zero (“At no time”) to five (“All of the time”). The raw total scores (0-25) are multiplied by four, producing final scores ranging from zero to 100, where higher scores indicate greater well-being. Prior work supports the reliability and validity of the WHO-5 (Topp et al., 2015). The baseline scores on the WHO-5 appeared internally consistent (*ω* = 0.90, 95% CI: 0.89, 0.91).

Emotion regulation

Emotion Regulation Scale (ERQ). The ERQ (Gross & John, 2003) is a ten-item self-report measure of individual differences in the use of two emotion regulation strategies: cognitive reappraisal (ERQ-CR; items 1, 3, 5, 6, 8, and 10) and expressive suppression (ERQ-ES; items 2, 4, 6, and 9). Prior work supports the reliability and validity of the ERQ in community samples (Preece et al., 2019, 2021). The ERQ items are rated on a seven-point Likert scale with responses ranging from one ("Strongly disagree") to seven ("Strongly agree"). We averaged item scores to produce final scores on the same metric of the original items (i.e., one to seven). The baseline scores on the ERQ-reappraisal (*ω* = 0.93, 95% CI: 0.92, 0.94) and ERQ-suppression (*ω* = 0.86, 95% CI: 0.84, 0.88) appeared to be internally consistent measures of emotion regulation.

Temperament

Ten Item Personality Inventory (TIPI). The TIPI assesses the "Big Five" personality traits (i.e., emotional stability, extraversion, openness to experience, conscientiousness, and agreeableness, Gosling et al. (2003)). The TIPI has 10 items (i.e, 2 per personality trait). They are rated on a 7-point Likert scale with higher scores indicating a greater endorsement of the personality traits. We averaged item scores to be on a 1-7 scale.

NLP Metrics

In addition to VADER (see Study 1 methods), we analyzed the text in the intervention using the Sentiment Analysis and Social Cognition Engine (SEANCE).

SEANCE is an NLP tool that uses an ensemble of algorithms, including VADER, to analyze sentiment along with other linguistic categories (Crossley et al., 2017). SEANCE outputs the NLP metrics associated to its constituent algorithms leading to anywhere from 250-3,000 different potential properties of text. The developers of SEANCE conducted a principal components analysis of all these metrics to identify a small number of components that could account for a large proportion of the variance across all the indices. We focus on the following SEANCE components: (1) a "Negative adjectives" component reflecting the use of negative descriptors (e.g., "terrible") including descriptions of negative emotions like disgust, anger, and sadness, (2) a composite of indices of "Positivity," which was created by adding three SEANCE component: positive adjectives, nouns and verbs (e.g., "smiling," "positive," "pleasant"), (3) a "Certainty" component indicating depiction of large quantities (e.g., "all"), exaggerated statements (e.g., "absolutely"), and conviction (e.g., "I am sure…"), and (4) an "Action" component indicating approach-related terms including descriptions of attempting actions (e.g. "try"), physical activity (e.g., "go"), and descriptive action verbs (e.g., "dancing")

Analysis plan

All analyses were conducted using the R programming language in the R Studio GUI. We relied on the *tidyverse* language for most data manipulation (Wickham et al., 2019).

To assess changes during the cognitive restructuring task (Aim A), we measure pre-post task changes in self-reported negative affect (VAS), sentiment (VADER compound scores), negative adjectives (SEANCE), positivity (SEANCE), action (SEANCE), and certainty (SEANCE). We present the changes as SMDs pre-post task. Next, we use a robust regression in *MASS* to regress changes on the VAS on pre- and post-restructuring task on: sentiment (VADER scores), Negative adjectives (SEANCE), Positivity (SEANCE), Action (SEANCE), and Certainty (SEANCE) while controlling for pre-task VAS.

To assess individual differences in emotion and emotion regulation in the cognitive restructuring task (Aim B), we used a canonical regression with the *R* packages *CCA* and *CCP* to correlate the NLP indices pre and post-cognitive restructuring (i.e., VADER sentiment and SEANCE Negative adjectives, Positivity, Action, and Certainty) with the self-reports obtained during the baseline period of the trial (i.e., PHQ-9 depression, GAD-7 anxiety, reappraisal and suppression on the ERQ, and personality in the TIPI). A canonical regression works by attempting to find multiple correlations between variable sets (i.e., the NLP metrics vs. the self-reported questionnaires). The procedure creates “variates” or synthetic predictors in the multivariate set. In these data, the variates would be linear combinations of the NLP metrics that correlate with linear combinations of the self-report variables. The variates can be conceptualized as factors in a factor analysis, (Tabachnick & Fidell, 2007), although it is more appropriate to consider factor analysis a special case of canonical correlation.

Aim C evaluated whether the NLP metrics predict depression and well-being at weeks 2 and week 8. Given the large number of NLP predictors, we used a machine-learning approach, least absolute shrinkage and selection operator (LASSO) with 10-fold cross-validation as implemented in *glmnet*, to identify predictors of post-treatment outcomes. LASSO shrinks regression coefficients to produce predictions that are more conservative, and thus more likely to generalize. Because LASSO can shrink regression coefficients to 0, it performs both variable selection and regularization.

Given that the NLP metrics from SEANCE do not have a natural interpretation (i.e., they are components from a principal components analysis), and that the self-reports and NLP metrics are on different scales, we computed Z scores for these variables and present all relations as standardized in the current sample.

Results

The sample consisted mostly of adults (*M* = 35.31, *SD* = 11.78), two thirds of whom identified as cisgender women (*n*=256, 66.14%). Most identified as Non-Hispanic White with a roughly equal representation of individuals who were Non-Hispanic Black (*n* = 21, 5.20%), Hispanic (*n*=27, 6.70%), or belonging to another racial-ethnic category (e.g., Asian, multiracial, *n*=44, 11.00%)

Aim 2a: NLP metrics as indicators of immediate emotion regulation

Individuals reported relatively high levels of distress (VAS) associated with the situations and automatic thoughts they recounted (M = 72.21, SD = 19.65). After the cognitive restructuring task, individuals reported lower levels of self-reported distress on the VAS (M = 51.65, SD = 21.54) suggesting that the cognitive restructuring task was associated with large reductions in negative affect (see Table 4). The NLP indices also suggest an improvement in the overall sentiment of the thoughts with a much lower overall VADER score post-task. The SEANCE metrics suggest that this change in sentiment is driven by a decreased Negative adjectives rather than an increase in Positivity. Interestingly, our analyses suggest an increase in the Action component but not a decrease in Certainty. The VADER scores post-task were the only statistically significant predictors of change in distress such that a more positive sentiment post-task was associated with more changes in self-reported distress (*β* = 0.12, 95% CI: 0.001, 0.25).

Table 4. Pre-post cognitive restructuring changes in self-reported distress (VAS) and NLP indices extracted from text data (N=387)

Metric	SMD	95%	CI
Negative affect (VAS)	-0.98	-1.20	-0.76
Affect (NLP: VADER)	0.60	0.39	0.81
Negative adjectives (NLP: SEANCE)	-0.52	-0.73	-0.32
Positivity (NLP: SEANCE)	0.13	-0.07	0.33
Action (NLP: SEANCE)	0.37	0.17	0.57
Certainty (NLP: SEANCE)	0.09	-0.11	0.29

SMD = Standardized Mean Differences, NLP = Natural Language Processing, VAS = Visual Analogue Scale, VADER = Valence Aware Dictionary for Sentiment Reasoning, Sentiment Analysis and Social Cognition Engine (SEANCE)

Aim 2b: NLP metrics as correlates of individual differences

Next, we used canonical correlation to regress the different NLP indices to measures of self-reported internalizing distress (PHQ-9, GAD-7, WHO-5, and ISI), emotion regulation (ERQ-CR and ERQ-ES), and temperament (extraversion, emotional stability, conscientiousness, agreeableness, and openness to experience) included in the trial. The canonical correlation analysis suggested the presence of one canonical dimension (*r*= 0.32, *p* = 0.03) but the test for more than one canonical dimension was not statistically significant (*p* = 0.26). The standardized canonical coefficients are presented in Table 5. From the self-reported screening questionnaires, the first dimension consisted of a strong loading from positive aspects of mental health including cognitive restructuring for emotion regulation (ERQ-CR: *r*=-0.56), extraversion (TIPI: *r* = 0.49), and emotional stability (TIPI: *r* = 0.38). For the NLP measures, the first dimension consisted of loadings indicating an initial response to the task that was less negative (SEANCE-Negative adjectives pre-task: *r* = -0.60) and less certain (SEANCE-Certain pre-task: *r* = -0.50) initial reaction to the task followed by a reduced certainty after the task (SEANCE-Certain post-task: *r* = -0.36).

Table 5. Canonical dimension representing the association between individual differences self-reports and NLP metrics from the cognitive restructuring task (N=387)

Metric	Dimension	β
Self-report	Depression (PHQ-9)	0.07
	Anxiety (GAD-7)	-0.08
	Well-being (WHO-5)	0.12
	Reappraisal (ERQ-CR)	0.44
	Suppression (ERQ-ES)	-0.14
	Emotional stability (TIPI-ES)	0.42
	Openness (TIPI-O)	-0.32
	Conscientiousness (TIPI-C)	-0.19
	Extraversion (TIPI-E)	0.48
	Agreeableness (TIPI-A)	-0.26
NLP metrics	Affect post (NLP: VADER)	0.27
	Negative post (NLP: SEANCE)	-0.21
	Positivity post (NLP: SEANCE)	-0.31
	Action post (NLP: SEANCE)	0.11
	Certainty post (NLP: SEANCE)	-0.32
	Affect pre (NLP: VADER)	-0.10
	Negative pre (NLP: SEANCE)	-0.53
	Positivity pre (NLP: SEANCE)	0.27
	Action pre (NLP: SEANCE)	0.32
	Certainty pre (NLP: SEANCE)	-0.58

PHQ = Patient Health Questionnaire, GAD-7 = Generalized Anxiety Disorder Scale 7, WHO-5 = WHO Well-being Index 5, ERQ = Emotion Regulation Questionnaire, TIPI = Ten-Item Personality Inventory, NLP = Natural Language Processing, VADER = Valence Aware Dictionary for Sentiment Reasoning, Sentiment Analysis and Social Cognition Engine (SEANCE)

Aim 2c: Prediction of symptom progression from NLP metrics

Results of the LASSO model are presented as standardized coefficients in Table 4. The strongest predictors of week 2 and week 8 depression (PHQ) and well-being (WHO-5) were the baseline scores on each of those measures. The next strongest predictor across most of the outcomes appeared to be the self-reported distress on the VAS after the cognitive restructuring task, although this effect was small. For the prediction of week 2 depression and well-being, the Certainty component post-task (SEANCE) was the strongest NLP predictor in the expected direction: use of more “certain” language was associated with higher depression and lower well-being, with the effect being substantially stronger for well-being than depression. The other NLP predictors of week 2 were mostly in the expected direction (e.g., more negativity in the initial thoughts was associated with greater depression at the week 2 follow-up) but with rather small effects (e.g., *β*s 0.01-0.05). The prediction of week 8 scores from the NLP metrics produces more mixed results. For the PHQ-9 scores, lower week-8 depression was associated with less positive sentiment (from both VADER and SEANCE) and less in the way of action words, but these effects were very small. In the prediction of week 8 WHO-5, in addition to the effects being small, these were also contrary to predictions. For example, higher week-8 well-being was predicted by more certainty language SEANCE post-restructuring task.

Table 6. Prediction of week 2 (n=311) and week 8 (n=299) depression (PHQ-9) and well-being (WHO-5) from NLP metrics pre- and post-cognitive restructuring

Variable	PHQ-9 (Week 2)	WHO-5 (Week 2)	PHQ-9 (Week 8)	WHO-5 (Week 8)
Intercept	0.00	0.00	-0.00	0.00
Distress pre-task (VAS)	0.00	0.00	0.00	0.00
Distress post-task (VAS)	0.06	-0.06	0.04	-0.11
Baseline severity	0.79	0.72	0.69	0.63
Sentiment post-task (VADER)	0.00	0.00	0.00	0.00
Negative component post-task (SEANCE)	0.00	0.03	0.00	0.06
Positive component post-task (SEANCE)	0.00	0.00	0.00	0.00
Action component post-task (SEANCE)	0.00	0.00	-0.00	0.00
Certainty component post-task (SEANCE)	0.03	-0.11	0.00	0.00
Sentiment pre-task (VADER)	0.00	0.00	0.00	0.00
Negative component pre-task (SEANCE)	0.02	0.00	0.00	0.00
Positive component pre-task (SEANCE)	-0.01	0.00	0.00	-0.03
Action component pre-task (SEANCE)	0.00	0.02	0.00	0.00
Certainty component pre-task (SEANCE)	0.00	0.06	0.00	0.05

Conclusions

Our Study 2 results support the utility of NLP metrics of negative and positive emotions, behavior (an action focus), and cognition (certainty) as helping to characterize individual differences in emotion and emotion regulation in psychopathology and support their use as capturing emotion regulation in the context of a CBT-based intervention. One implication of our findings is that it may be possible to embed NLP-type metrics in the context of CBT interventions to provide real-time feedback about the success of different activities (e.g., the success in cognitive restructuring). However, the data do not provide strong support for the role of NLP metrics, above and beyond self-report in predicting symptom change following treatment. While several NLP metrics predicted symptom changes at the 2-week follow-up post-intervention, their effects were small and then became smaller and less consistent when predicting 8-week depression scores. Thus, more work is needed to uncover whether and how NLP metrics relate to symptoms changes.

Discussion

While our results are exciting and offer novel insights into how psychopathology and emotions manifest in text, several limitations are worth considering. First, aside from the within-person changes in the cognitive restructuring task, most of the effects that we report are relatively small. This is a known issue in correlating outcomes that are obtained from different methods (i.e., self-report vs. NLP). Moreover, in several analyses we report predictive validity of NLP metrics above and beyond self-report. We used pre-existing NLP tool where perhaps it could be argued that developing specific NLP algorithms specific to this task (e.g., that are trained to detect changes in distress during the task) may uncover larger effects. Finally, we had a wealth of self-report, text data, and associated NLP metrics to choose from in these analyses. In such situations, there is a high number of “researcher degrees of freedom” that may influence the results (Wicherts et al., 2016). We tried to balance using theory (e.g., choosing different SEANCE metrics) and data-driven approaches (e.g., using dimension reduction techniques and machine learning) to minimize the possibility of spurious findings. Nonetheless, spurious findings are still a possibility.

We have illustrated the triangulation of self-reported measures of psychopathology and NLP metrics using data derived from passively-acquired observations of social media activity (CONCEALED FOR PEER REVIEW) along with data from a clinical trial (CONCEALED FOR PEER REVIEW). These studies suggest that emotional dynamics can be passively captured and are differentially associated with dimensions of psychopathology (Study 1, Study 2 Aim B).

Moreover, our results suggest that NLP can also be used to study emotion regulation within a treatment context (Study 2 Aim A), above and beyond self-report, and, to a more limited extent, post-treatment psychopathology (Study 2 Aim C).

Data transparency

Please note that the data for study 1 have only been used twice in Lorenzo-Luaces et al. (2022) and Rutter et al. (2023). Lorenzo-Luaces et al. (2022) examined differences between individuals in the cohorts by the type of social media account they provided (i.e., not at all, a valid account, a bot-like account). In the current study, we only use individuals with valid Twitter accounts. Rutter et al. (2023) was a re-analysis of self-report data with a focus on individuals "self-diagnosing" mental disorders. The data for Study 2 have only been used once: Lorenzo-Luaces & Howard (2022). That paper reported the main results of the clinical trial. In this study, we only use the treatment condition, with a focus on the text data from the intervention.

References

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529–542.

Bathina, K. C., Thij, M. ten, Lorenzo-Luaces, L., Rutter, L. A., & Bollen, J. (2021). Individuals with depression express more distorted thinking on social media. *Nature Human Behaviour*, *5*, 458–466.

Beck, A. T., & Haigh, E. A. (2014). Advances in cognitive theory and therapy: The generic cognitive model. *Annual Review of Clinical Psychology*, *10*, 1–24.

Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). Happiness is assortative in online social networks. *Artificial Life*, *17*(3), 237–251.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology*, *63*(9).

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*, *49*, 803–821.

Fried, E. I., Borkulo, C. D. van, Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, *28*(11), 1354.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Rersonality*, *37*(6), 504–528.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348.

Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion*, *13*(1), 1–6. https://doi.org/10.1037/a0030262

Holtzman, N. S. et al. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, *68*, 63–68.

Houben, M., Noortgate, W. V. D., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 216–225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Jazaieri, H., Urry, H. L., & Gross, J. J. (2013). Affective disturbance and psychopathology: An emotion regulation perspective. *Journal of Experimental Psychopathology*, *4*(5), 584–599.

Jenkins, B. N., Hunter, J. F., Cross, M. P., Acevedo, A. M., & Pressman, S. D. (2018). When is affect variability bad for health? The association between affect variability and immune response to the influenza vaccination. *Journal of Psychosomatic Research*, *104*, 41–47. https://doi.org/10.1016/j.jpsychores.2017.11.002

JH Balsters, M., J Krahmer, E., GJ Swerts, M., & JJM Vingerhoets, A. (2012). Verbal and nonverbal correlates for depression: A review. *Current Psychiatry Reviews*, *8*(3), 227–234.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, *120*(2), 263–286.

Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Hobbs, K., Reininghaus, U., Slade, T., et al. (2020). Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, *19*(2), 151–172.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., et al. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477.

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., et al. (2021). Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, *20*(2), 171–193.

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion*, *7*(2), 262.

Leger, K. A., Charles, S. T., & Fingerman, K. L. (2019). Affect variability and sleep: Emotional ups and downs are related to a poorer nights rest. *Journal of Psychosomatic Research, 124*, 109758. https://doi.org/10.1016/j.jpsychores.2019.109758

LeMoult, J., Kircanski, K., Prasad, G., & Gotlib, I. H. (2017). Negative self-referential processing predicts the recurrence of major depressive episodes. *Clinical Psychological Science, 5*(1), 174–181.

Liu, T., Meyerhoff, J., Eichstaedt, J. C., Karr, C. J., Kaiser, S. M., Kording, K. P., Mohr, D. C., & Ungar, L. H. (2022). The relationship between text message sentiment and self-reported depression. *Journal of Affective Disorders*, *302*, 7–14.

Lorenzo-Luaces, L. (2015). Heterogeneity in the prognosis of major depression: From the common cold to a highly debilitating and recurrent illness. *Epidemiology and Psychiatric Sciences*, *24*(6), 466–472.

Lorenzo-Luaces, L., Buss, J. F., & Fried, E. I. (2021). Heterogeneity in major depression and its melancholic and atypical specifiers: A secondary analysis of STAR* D. *BMC Psychiatry*, *21*, 1–11.

Lorenzo-Luaces, L., German, R. E., & DeRubeis, R. J. (2015). It's complicated: The relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. *Clinical Psychology Review*, *41*, 3–15.

Lorenzo-Luaces, L., & Howard, J. (2022). *Efficacy of a single session intervention for depression in online workers: A randomized controlled trial with transdiagnostic mental health outcomes*.

Lorenzo-Luaces, L., Howard, J., Edinger, A., Yan, H. Y., Rutter, L. A., Valdez, D., Bollen, J., et al. (2022). Sociodemographics and transdiagnostic mental health symptoms in SOCIAL (Studies of Online Cohorts for Internalizing Symptoms and Language) I and II: Cross-sectional survey and botometer analysis. *JMIR Formative Research*, *6*(10), e39324.

Lorenzo-Luaces, L., Keefe, J. R., & DeRubeis, R. J. (2016). Cognitive-behavioral therapy: Nature and relation to non-cognitive behavioral therapy. *Behavior Therapy*, *47*(6), 785–803.

Lorenzo-Luaces, L., Peipert, A., De Jesus Romero, R., Rutter, L. A., & Rodriguez-Quintana, N. (2021). Personalized medicine and cognitive behavioral therapies for depression: Small effects, big problems, and bigger data. *International Journal of Cognitive Therapy*, *14*, 59–85.

Luchman, J. N., Bergstrom, J., & Krulikowski, C. (2014). A motives framework of social media website use: A survey of young americans. *Computers in Human Behavior*, *38*, 136–141.

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., et al. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: A systematic analysis for the global burden of disease study 2010. *The Lancet*, *380*(9859), 2197–2223.

Perrin, A. (2015). Social media usage. *Pew Research Center*, *125*, 52–68.

Preece, D. A., Becerra, R., Hasking, P., McEvoy, P. M., Boyes, M., Sauer-Zavala, S., Chen, W., & Gross, J. J. (2021). The emotion regulation questionnaire: Psychometric properties and relations with affective symptoms in a United States general community sample. *Journal of Affective Disorders*, *284*, 27–30.

Preece, D. A., Becerra, R., Robinson, K., & Gross, J. J. (2019). The emotion regulation questionnaire: Psychometric properties in general community samples. *Journal of Personality Assessment*.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Riley, T. N., Thompson, H. M., Howard, J., Lorenzo-Luaces, L., & Rutter, L. A. (2022). Seeking connectedness through social media use: Associations with adolescent empathic understanding and perspective-taking. *Current Psychology*, 1–13.

RStudio Team. (2020). *RStudio: Integrated development environment for r*. RStudio, PBC. http://www.rstudio.com/

Rutter, L. A., Howard, J., Lakhan, P., Valdez, D., Bollen, J., & Lorenzo-Luaces, L. (2023). "I haven't been diagnosed, but i should be"—insight into self-diagnoses of common mental health disorders: Cross-sectional study. *JMIR Formative Research*, *7*(1), e39206.

Scott, L. N., Victor, S. E., Kaufman, E. A., Beeney, J. E., Byrd, A. L., Vine, V., Pilkonis, P. A., & Stepp, S. D. (2020). Affective dynamics across internalizing and externalizing dimensions of psychopathology. *Clinical Psychological Science*, *8*(3), 412–427. https://doi.org/10.1177/2167702619898802

Sperry, S. H., Walsh, M. A., & Kwapil, T. R. (2020). Emotion dynamics concurrently and prospectively predict mood psychopathology. *Journal of Affective Disorders*, *261*, 67–75.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, *84*(3), 167–176.

Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J. S., & DeRubeis, R. J. (2021). Promoting graduate student mental health during COVID-19: Acceptability, feasibility, and perceived utility of an online single-session intervention. *Frontiers in Psychology*, *12*, 569785.

Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J., & DeRubeis, R. J. (2022). Promoting mental health with an inexpensive, online single-session intervention during the COVID-19 pandemic: A randomized controlled trial. *Manuscript Submitted for Publication*.

Watson, D., Levin-Aspenson, H. F., Waszczuk, M. A., Conway, C. C., Dalgleish, T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Hobbs, K. A., et al. (2022). Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): III. Emotional dysfunction superspectrum. *World Psychiatry*, *21*(1), 26–54.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775.

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 1832.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D.,...R. F., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yang, K.-C., Ferrara, E., & Menczer, F. (2022). Botometer 101: Social bot practicum for computational social scientists. *arXiv Preprint arXiv:2201.01608*.