

Naïve Observers' Perceptions of Hospitalized and Community Adolescents' Depressive Symptoms

Marta Tironi

University of Genoa

Abby Rozenberg

Morgan McLoughlin

Jane Shkel

University of California, Irvine

Breana R. Cervantes

University of Houston

Andres De Los Reyes

University of Maryland

Carla Sharp

University of Houston

Jessica L. Borelli

University of California, Irvine

Adolescent depression significantly impacts long-term functioning and well-being, often going undetected and untreated, highlighting the need for innovative detection

methods. This study examined whether unfamiliar, naïve observers could detect depressive symptoms in adolescents, relying solely on behavioral residue (language transcribed from a clinical interview). Identifying whether naïve

The findings from this paper have not been published previously and this manuscript is not under consideration elsewhere. The data were derived from a larger study (Ibraheim et al., 2017; Venta et al., 2012); however, for the current investigation, we rely solely on the Child Attachment Interview (CAI; Shmueli-Goetz et al., 2008) and adolescents' reports of their depressive symptoms (Beck Depression Inventory-II; Beck et al., 1996). Naïve observers reviewed the CAI transcripts with no other knowledge of the participants and provided their assessment of participants' depressive symptoms on the BDI. The resultant dataset is unique from what has previously been published as we are interested in the associations between observers' ratings and target participants' ratings.

We did not have any grant funding for this particular study, but the broader study (from which the adolescents participant data were drawn) was funded by the Child and Family Program of the Menninger Clinic in Venta et al. (2012), and by the McNair Family Foundation and the Duncan Foundation in Ibraheim et al. (2017).

The authors declare no conflict of interest.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. Data has not been shared on a repository, but they can be requested to the corresponding author.

Ethics statement: All procedures were conducted in accordance with the APA ethical guidelines, and the manuscript is not posted in any format online. This study was approved by the Institutional Review Board of the University of Houston (HA IRB#: 14238-02, NHA IRB# 13469-02), and later by a reliance from the IRB at UC Irvine. Participants were treated in accordance with the ethical principles stated in the Declaration of Helsinki and given their written consent before the study participation.

Address correspondence to Jessica L. Borelli, Ph.D., University of California Irvine, Department of Psychological Science, 4201 Social & Behavioral Sciences Gateway, Irvine, CA 92697.

0005-7894/© 2025 Association for Behavioral and Cognitive Therapies. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>); email: jessica.borelli@uci.edu.

observers, untrained in detecting psychological symptoms, can utilize verbal output to accurately detect depressive symptoms yields scientific insight that could be expanded upon to support prevention efforts. The study involved 200 adolescent participants (100 hospitalized, 100 community; $M_{\text{age}} = 15.31$, henceforth referred to as targets). They reported on their depressive symptoms and completed the Child Attachment Interview (CAI; Shmueli-Goetz et al., 2008). Naïve observers ($N = 52$ college students split in two rounds), unaware of all other information about targets, read transcripts of targets' CAI interviews; using solely that information, they rated targets' depressive symptoms. Observers also assessed their confidence in their evaluation of targets' symptoms. Results showed a positive association between observers' ratings and targets' self-reported depressive symptoms (Round 1 Cohen's $d = 1.25$, Round 2 Cohen's $d = 1.42$) and this association remained significant controlling for linguistic categories connected to depression detected by computer analysis. Observers' ratings discriminated between hospitalized and community adolescents (Round 1 Cohen's $d = 1.43$, Round 2 Cohen's $d = 1.37$), though confidence in their ratings was not associated with targets' ratings or group (community or hospitalized). Observers can reliably and accurately identify depressive symptoms in adolescents using attachment interviews, over and above computer analysis alone. This work suggests that naïve observers contribute unique insight into the detection of adolescent depressive symptoms.

Keywords: depressive symptoms; person perception; attachment; adolescence; naïve observer

ADOLESCENCE is a critical developmental period in which individuals are particularly susceptible to developing mental health problems, including depression (Leussis & Andersen, 2008). The prevalence of major depressive disorder increases rapidly with the onset of adolescence, especially among females (Lu, 2019; NHS, 2021), with estimates ranging up to 35% (Shorey et al., 2022; UK Office of National Statistics, 2020). While early intervention for mood disorders is effective when implemented, adolescents in the U.S. experiencing these symptoms are less likely to be identified by others and often go longer before receiving appropriate treatment (Costello et al., 2006). Barriers to accurately identifying adolescents' depressive symptoms pose a significant issue, as undetected symptoms tend to worsen over time (Stuckey et al., 2021), and recent reports suggest that research and policy should concentrate on efforts to target teenage depression symptoms before they

worsen (e.g., Shorey et al., 2022). In addition to the risk that depression poses on its own, depressive symptoms are also present at subclinical levels in many different psychopathologies, such as obsessive-compulsive disorder (Rickelt et al., 2016) or substance abuse disorder (Fleming et al., 2019), and also co-occur at high rates in anxiety disorders (Schubert et al., 2017). Depression worsens prognosis for these conditions (e.g., Melton et al., 2016; Rickelt et al., 2016), perhaps because depressive symptoms may make it more difficult to adhere to treatment regimens (Mojtabai et al., 2002).

Depression is commonly defined based on subjective perceptions of distress and emotional states. The internal nature of the disorder may contribute to difficulties in accurate detection of this condition in youth, particularly by the adults involved in their lives (e.g., parents, teachers). Particularly during adolescence, a developmental period characterized by a heightened sense of what information one should or should not disclose to others, depressive symptoms may be masked, making detection more difficult (e.g., Smetana, 2008).

However, adolescents' verbal behavior can reveal important cues about their psychological states, including their depressive symptoms. For instance, irritability, low mood, and lethargy of depression may be able to be detected via verbal behavior. Given that a large proportion of adolescent communication occurs via verbal behavior (via texting or social media, e.g., Lenhart et al., 2010, but also via social interaction), efforts to detect depression through verbal, text-based behavior hold promise (Underwood et al., 2015). Indeed, adolescents' speech during family interactions (Low et al., 2010) as well as their spontaneous speech, as in an interview or conversation (Alghowinem et al., 2013), are associated with depressive symptoms. These encouraging findings demonstrate that depression can be identified via automatic data processing mechanisms (e.g., computer-based linguistic and speech analysis; see, for example, Funkhouser et al., 2023). What remains unknown is whether the same detection process can occur by human observers in adolescents' natural environments. Given that adolescents' speech is not typically recorded and analyzed by computers, identifying whether other people can detect depressive symptoms is of great interest. In the current study, we examine whether outside observers can detect signs of depressive symptoms in adolescents' verbal behavior and whether this is feasible at least in principle.

PERSON PERCEPTION METHODOLOGY: RELEVANCE FOR ADOLESCENT DEPRESSION DETECTION

The Operations Triad Model (De Los Reyes et al., 2013, 2023) contends that it is critical to gather and analyze data from multiple sources to understand youth's experiences. De Los Reyes argues that observations from people who know and are unknown to a youth provide unique information regarding a youth's experiences. Youth's behavior varies within and across different social settings (Achenbach et al., 2002; Whitcomb & Merrell, 2013). When it comes to understanding adolescents, seeking knowledge from sources other than parents, such as someone unfamiliar with the youth (e.g., from a "naïve observer," henceforth referred to as *observers*), can augment the assessment process (Makol et al., 2020) and provide an outside perspective on the youth's symptoms (e.g., adolescent social anxiety; Deros et al., 2018).

The person perception (PP) methodology, used by social psychologists to examine how individuals process information about and form impressions of others, evaluates whether people can detect traits of unfamiliar others based on small behavioral samples. These interactions are thought to approximate everyday "first impressions" (e.g., Gilbert, 1998). PP studies reveal humans' impressive accuracy in judging others based on limited information. Using these narrow behavioral samples (e.g., content from social media, speech), observers can form accurate inferences about a stranger's personality traits (Gosling et al., 2002; Kenny & Albright, 1987), attachment to a former partner (Borelli et al., 2019), or propensity for maladaptive behaviors (e.g., aggression/intimate partner violence; Russo & Borelli, 2023). Such accurate judgments are thought to be linked to the "behavioral residue" (Gosling et al., 2002) provided by even limited informational snapshots of a "target", i.e., the person being observed. More recent work has also examined language used by the target as a specific and highly informative form of behavioral residue (Borelli et al., 2018, 2019) that provides discrete cues regarding internal emotional states. Psychologists use this collateral information from people without mental health training to assess psychosocial functioning, including impairments, and to form insights into their symptoms (Charamut et al., 2022; Rezeppa et al., 2021).

Two key areas in PP include the accuracy and confidence of observers' ratings of targets' characteristics. Accuracy is operationalized as self-other agreement (Bernieri et al., 1994; Borelli et al., 2019; Russo & Borelli, 2023)—that is, the degree

to which observers' and targets' ratings are positively associated. Accuracy can be affected and improved by a range of factors, including characteristics of the observer (e.g., Forgas & Bower, 1987), the target (Montepare & Zebrowitz, 1998), and the quantity and quality of information provided to the observer (Biesanz et al., 2007). According to this framework, the visibility of traits, or how clearly observable others' personality traits and behaviors are (Watson et al., 2000), and the quality or quantity of the information they can rely on, can positively influence the accuracy of detection processes (Biesanz et al., 2007; Funder, 2018). Trait visibility may be particularly relevant to detecting internalizing problems such as depression, which are thought to be less detectable due to their "private" nature (Rodriguez et al., 2010). The opportunity to assess and observe such traits may also depend on the social environment and the context in which they are observed (e.g., degree of familiarity; Charamut et al., 2022).

Confidence refers to the level of certainty that observers have regarding their perceptions of targets' attributes. Prior PP research has found that observers' confidence is uncorrelated with accuracy (Waggoner et al., 2009). Nevertheless, confidence can have implications for real-world social and behavioral outcomes, leading to more competent behavior among providers (Fry et al., 2015) or, on the contrary, less accurate decision-making (Soll & Larrick, 2009). In the case of adolescents' mental health, greater confidence that a teenager is depressed (whether accurate or not) may lead to prosocial behavior and greater care for the affected individual both emotionally, by enhancing empathy and active listening, and operationally, by actively seeking help. Alternatively, greater confidence could lead to more stigmatizing and reflexively negative behavioral responses, with damaging effects for the targeted individuals. Thus, observers' confidence is relevant to understanding how it impacts their evaluations of depressive symptoms.

NAÏVE OBSERVERS' PERCEPTIONS ABOVE AND BEYOND LINGUISTIC CUES

Depression can be consistently detected by observers when they are presented with written outputs (i.e., online blogs or personal diaries; Rodriguez et al., 2010), while it has only been inconsistently detected from audio samples drawn from daily life and then presented to observers (Mehl et al., 2006). Written outputs of speech provide linguistic cues, which may be one important source of these perceptual judgments of depressive symptoms.

Moreover, as one of the most common forms of human communication, word use is a good indicator with which to begin to determine psychopathological risk in the PP paradigm.

Linguistic analyses of these texts have isolated specific word categories that are associated with depressive symptoms. For instance, negative emotion words (e.g., *hate, fear, tired*), cognitive mechanism words or cognitive processes (e.g., *but, if, know*), first-person singular and other self-focused pronouns (e.g., *I, me*), and past-tense verbs are associated with depressed mood and have all been implicated for their association with depressive or internalizing symptoms among adults (Cohn et al., 2004; Lyons et al., 2018; Rude et al., 2004). Moreover, using a text analysis of several online forum groups, Al-Mosaiwi and Johnstone (2018) found that adults with depression are more likely to use absolutist or all-or-none words (i.e., words that suggest magnitude or probability without nuances; e.g., *always, totally, entire*). Personal pronouns and absolutist/all-or-none words are both considered to be function words, and as such are less likely to be consciously controlled (Al-Mosaiwi & Johnstone, 2018; Pennebaker et al., 2003). Although little work has examined depression detection using linguistic analysis with adolescents, one study found that adolescents who used more first-person singular pronouns in smartphone communication were linked to increased chances of reaching the major depressive disorder threshold (Funkhouser et al., 2023).

Given their prior links with depressive symptoms, as well as the fact that some of them are not consciously controlled, these word categories (i.e., negative emotions, cognitive processes, personal pronouns, focus on the past, and absolutist/all-or-none words) may be detectable by computer programs. In a real-world scenario, language taken from written text (i.e., text messages, social media postings or school assignments) or speech samples (i.e., informal conversations and interviews within specific contexts, such as schools, therapy, hospital) can be analyzed to detect mental health functioning, with specific linguistic cues serving as proxies for depression.

CURRENT STUDY

This study, which was not preregistered, expands on prior PP research on depressive symptoms in adults by investigating observers' agreement, accuracy (relative to adolescent self-reports), and confidence in detecting depressive symptoms in hospitalized adolescents (HA) from an inpatient psychiatric clinic and adolescents from the

community, henceforth defined as community adolescents (CA) (*targets*). The purpose of this research is to provide a preliminary demonstration regarding the validity of this approach, generating future work regarding its potential practical implications. Specifically, this study is poised to test whether observers with no training related to PP, psychopathological symptom detection, or mental health are able to detect depressive symptoms using relevant linguistic information, such as the content derived from an interview about attachment relationships (the CAI; Shmueli-Goetz et al., 2008). The use of an emotionally evocative interview (CAI) may assist observers in the detection of a predominantly, internally manifested mental health condition, therefore identifying depression risk (Biesanz & Wallace, 2020). Our work is based on recent work using an evocative behavioral task that provided support for this approach using untrained observers (Follet et al., 2023; Rezeppa et al., 2021). This design allowed for parallel assessment of observer accuracy in depression, as well as examination of the role of symptom severity (using the proxy measure of the group: hospitalized vs community) in observers' ratings. We examine whether naïve observers' prediction is significant above and beyond the automatic linguistic analysis of the text itself.

In designing this study, we considered the most appropriate sample in which to evaluate our research questions, guided by two overarching goals. First, we thought it was necessary to select two groups that would be distinguishable in levels of symptom severity. Given that the prevalence of depression has increased within community samples of adolescents within the last decade (e.g., Lu, 2019), with up to 35% of adolescents within community samples having clinically significant depression (Shorey et al., 2022; UK Office of National Statistics, 2020), we thought a sample of HA would be an optimal higher acuity comparison sample. Second, we sought to examine whether observers could detect higher levels of depression as in the real world, where they present with diagnostic comorbidities (Demyttenaere & Heirman, 2020; Rohde, 2009). By focusing on HA (as opposed to a sample of adolescents exclusively diagnosed with major depressive disorder), our study enables detection of depressive symptoms across diverse clinical presentations, enhancing the generalizability of the findings.

We explored four aims. With Aim 1 we evaluated whether observers were internally consistent in their perception of both HA and CA targets' symptoms. Moderate or greater levels of reliability (i.e., Intraclass Correlation Coefficient [ICC]

higher than .50; Koo & Li, 2016) in their depression scores in both groups of adolescents (H1) would suggest that the raters at least moderately agreed with one another. With Aim 2 we assessed the degree of observers' accuracy in their perceptions of targets' depression. We hypothesized that observers' ratings of targets' depression would demonstrate moderate convergence with targets' self-ratings of their depressive symptomatology, operationalized as correlations of .40 and higher (cf. Watson et al., 2000) (H2). Aim 3 involved examining whether the linguistic categories as measured from the CAI transcripts that have previously been identified as related to depressive symptoms in the research literature (i.e., self-personal pronouns, negative emotions, all-or-none/absolutist words, focus on the past, and cognitive processes) predict targets' ratings, and whether including NOs improves the predictive power of the regression equation above and beyond linguistic analysis. This aim enables us to explore the incremental contribution of higher resource assessment approaches, such as utilizing groups of raters, over and above computer analysis alone. With Aim 4 we examined the magnitude of observers' depression ratings of targets' depressive symptoms as a function of the clinical group, predicting that observers' depression ratings would be higher in HA as compared to CA (H4a). We also predicted that observers' accuracy would be moderated by group, such that the convergence between observers' and targets' ratings (i.e., accuracy) would be higher in HA than in CA (H4b). These two predictions (H4a and H4b) were based on the idea that, in HA, symptom presentation would be stronger and more apparent (e.g., Thompson et al., 2020), leading to easier detection due to the visibility of traits. This would lead to higher rating values and accuracy. Finally, with Aim 5 we examined whether observers' confidence in their perceptions of targets' depression was positively associated with the targets' own reports of their symptomatology (H5a) and whether the group was associated with confidence (CA vs HA). Targets with more severe symptoms were expected to provide more numerous and reliable linguistic cues about their mood states, which may indirectly allow observers to have greater trust in their own evaluations.

Method

PARTICIPANTS

Target Participants

Target data were derived from two larger studies of adolescents between the ages of 12 to 17

(Ibraheim et al., 2017; Venta et al., 2012). Prior to data collection, the study protocols were approved by the Institutional Review Board of the University of Houston (for CA: IRB# 14238-02, for HA: IRB# 13469-02). Materials and analysis code for this study are available by emailing the corresponding author. From 2009 to 2014, two groups of adolescents between the ages of 12 and 17 years living in a large metropolitan city in the Southwestern United States were recruited to participate: (1) adolescents receiving treatment from an inpatient psychiatric clinic (HA) and (2) an age-matched sample recruited from the community free of any diagnosis (CA) (see Table 1 for demographics information). HA participants were in a residential treatment facility offering psychodynamic treatment for social-cognitive growth, emotional regulation, maladaptive behaviors, and family relations. HA received a mental health assessment on admission and interviews were collected primarily occurring within 3 days of their hospital admissions. The average length of hospitalization was 30.90 days ($SD = 9.94$). Based on the CDISC-IV (the computerized Diagnostic Interview Schedule for Children based on DSM-IV; Shaffer et al., 2000) conducted with adolescents at admission, 51.1% of the sample met criteria for a depressive disorder, 52.8% for an anxiety disorder, 3.3% for a bipolar disorder, 33.3% for an externalizing disorder, 8.9% for an eating disorder, 5.5% borderline personality disorder, and 63.7% for emerging personality disorders. The sample had considerable diagnostic comorbidity.

CA were recruited from local schools and community organizations. Eligibility criteria included being an adolescent between 12 to 17 years old, fluent in English, without a documented diagnosis of autism spectrum disorder, schizophrenia or other psychotic disorder, and IQ over 70 as documented by parents' report and by patient history in current medical records. Adolescents with comorbid psychopathology and previous diagnosis and treatment for speech and language impairments were not excluded from participation.

Two hundred ($M_{age} = 15.31$; $SD_{age} = 1.27$) participants were selected from the larger data repositories based on completion of the primary study variables of interest in the current study (CAI and BDI-II). The sample was majority female (71.5%). For the purpose of this study, we grouped participants into one of two racial categories (White, $N = 43.85\%$ or Non-White, $N = 56.15\%$) due to the limited heterogeneity especially within the HA sample.

Table 1
Sociodemographic Characteristics of Targets

	HA		CA	
	<i>n</i>	%	<i>n</i>	%
Gender				
female	69	69.00	74	74.40
male	31	31.00	25	25.30
Diagnosis				
depressive disorder	51	51.10		
anxiety disorder	53	52.80		
bipolar disorder	3	3.30		
externalizing disorder	33	33.30		
eating disorder	9	8.90		
Race/ethnicity				
white	72	80.90	10	10.20
non-white	17	19.10	88	89.80
Parents' highest education				
some high school	0	0.00	5	5.00
high school diploma	2	2.00	12	12.00
some college	15	15.00	15	15.00
bachelor's degree	43	43.00	33	33.00
master's degree	18	18.00	17	17.00
doctoral degree	13	13.00	3	3.00
other	0	0.00	2	2.00
not answered	9	9.00	13	13.00
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Age (years)	15.40	1.40	15.20	1.16
Length of hospitalization (days)	30.90	9.94		

Target Procedure

After completing informed consent (parents) and assent (adolescents), participants from the two groups completed the CAI and the BDI-II (within 3 days of admission). Video-recorded CAIs were transcribed verbatim. Interviewer verbalizations were removed from the transcripts so only the adolescents' statements remained. Compensation was offered to only the CA sample.

Target Measures

Behavioral Sample. Both the HA and CA completed the CAI (Shmueli-Goetz et al. 2008), a semistructured interview that examines children's representations of their current relationships with their primary caregivers to measure their attachment, using both a dimensional (11 subscales) and a categorical approach (e.g., secure, insecure-dismissing, insecure-preoccupied). The CAI's 19 questions probe the adolescent's self-representation, those of their primary caregivers, as well as moments of conflict, discomfort, sickness,

hurt, separation, and loss. Consistent with interview protocol, the interview is video-recorded and transcribed verbatim. In previous studies the CAI proved useful in predicting self-reported internalizing symptoms (Borelli et al., 2010; Venta et al., 2015) and has high interrater reliability (Shmueli-Goetz et al. 2008; Venta et al., 2015).

Depressive Symptoms. Targets completed the Beck Depression Inventory, 2nd ed. (BDI-II; Beck et al., 1996), a 21-item self-report survey using Likert-type scaling. The BDI-II requires respondents to rate items probing symptoms of sadness, loss of interest, and worthlessness on a 4-point scale (from 0 = *no depressive symptom*, e.g., "I do not feel sad" to 3 = *maximum presence of depressive symptom*, e.g., "I am so sad and unhappy that I can't stand it"), with higher scores implying greater depressive symptoms. In this study, analysis of the targets' total score yielded good internal consistency of $\alpha = .93$ for HA and $\alpha = .90$ for CA. The efficacy of this measure in assessing adolescent depressive symptoms has been demonstrated in prior studies (e.g.,

Rausch et al., 2017; Rezeppa et al., 2021). The BDI-II has demonstrated validity and reliability with youth as young as 12, as well as with inpatients and nonclinical adolescents (Kumar et al., 2002; Osman et al., 2004, 2008).

Naïve Observer Participants

The study design has built-in replication: two independent samples of observers were recruited. Round 1 consisted of $N = 16$ observers (50% female, $M_{age} = 21.3$, $SD_{age} = 1.45$, 37.5% White, 43.8% prefer not to say, 18.8% Asian). Round 2 consisted of $N = 36$ observers (80.6% female, $M_{age} = 21.8$, $SD_{age} = 1.77$, 36.1% White, 36.1% prefer not to say, 22.2% Asian, 5.6% Others). Observers were college students enrolled in a research course with no prior training in PP or mental health nor in research. Observers' eligibility criteria included: over 18 years of age, fluency in speaking, reading, and writing in English, and attending college. We engaged in two rounds of coding: Round 1 was largely exploratory, enabling us to focus only on observers' agreement (*between observers*) and accuracy (*with target ratings*). Round 2 enabled researchers to assess agreement, accuracy of observers with targets, observers' confidence in their ratings, and the specific linguistic cues used to generate their ratings.

Naïve Observer Procedure

The researcher verbally explained the study protocol to each observer. First, each observer was presented with a random set of deidentified transcripts. Every CAI transcript was rated by 3 observers (randomly assigned), allowing for the examination of interrater reliability amongst the raters (see Supplemental Documents Section 1, "Naïve Observers' Rating Procedures", for details regarding the recruitment and coding procedures of the observers). Each rater was given 10 to 13 transcripts weekly for up to 15 weeks to complete the interviews. Transcripts included adolescents' answers to the CAI questions after eliminating all interviewer comments. Observers were informed only that: (1) the transcripts were a speaker completing an attachment interview; (2) the interviewees were adolescents between ages 12 to 17; (3) some adolescents were hospitalized for inpatient psychiatric treatment at the time of the interview, and some were not. Observers were not informed as to which targets were and were not hospitalized. Observers had no background in psychology, and they were told that reading each transcript would help them gain an understanding of the adolescent's perspective. Solely relying on the information gleaned from the transcript, observers were assigned

to complete the BDI-II as if they were the adolescent speaking in the transcript. They completed the BDI-II in the first person because that is how the standardized questionnaire is designed and validated (e.g., item 1: "*I am depressed all the time*"); administering it in this way did not entail modifying the questionnaire. Furthermore, the standard PP framework suggests that identification with the targets can serve as reinforcement for a more accurate assessment (e.g., Hurtig & Pichevin, 1990, with sex identification). In Round 2, observers were given identical instructions as in Round 1 but, in addition to completing the procedures of Round 1 observers, they also completed a measure to assess their confidence in their ratings (see Figure 1).

Naïve Observer Measures

Depressive Symptoms. Observers completed the same depression measure completed by the targets (i.e., BDI-II), to provide their perceptions of the targets' depressive symptoms. Observers' internal consistency in Round 1 yielded a score of $\alpha = .97$ for HA and a score of $\alpha = .96$ for CA; in Round 2 the α score was .93 for HA and .95 for CA.

Naïve Observers' Confidence. After reading each CAI transcript and completing a BDI-II for the target, observers rated their degree of confidence by responding to a statement modeled after our prior work (see Borelli et al., 2019), i.e., "*I am confident that my ratings accurately captured this participant's depression (or lack of depression)*", on a 5-point Likert scale from 1 (*no confidence*) to 5 (*very high confidence*).

Automatic Linguistic Analysis. In making their ratings of targets, the observers solely relied on the targets' transcribed verbal responses to the CAI. The CAI tends to elicit emotionally charged personal information, which may have an impact on target language patterns (Goldberg, 2014). To understand whether observers' evaluation may be associated with targets' rating above and beyond specific linguistic cues from the CAI transcripts detected by automatic linguistic analysis, we created a text file for each target's transcript and used them as inputs for the linguistic analysis. We analyzed these transcripts with the Linguistic Inquiry and Word Count (LIWC-22; Pennebaker et al., 2007), using the following categories based on literature: first-person singular pronouns (e.g., *I, we, me*), negative emotions (e.g., *bad, hate, tired*), focus on the past (e.g., *was, had, were*), all-or-none (i.e., absolutist words such as *all, never, always*), and cognitive processes (e.g., *but, not, if, know*). Thus, each target had a single score on each of these linguistic variables.

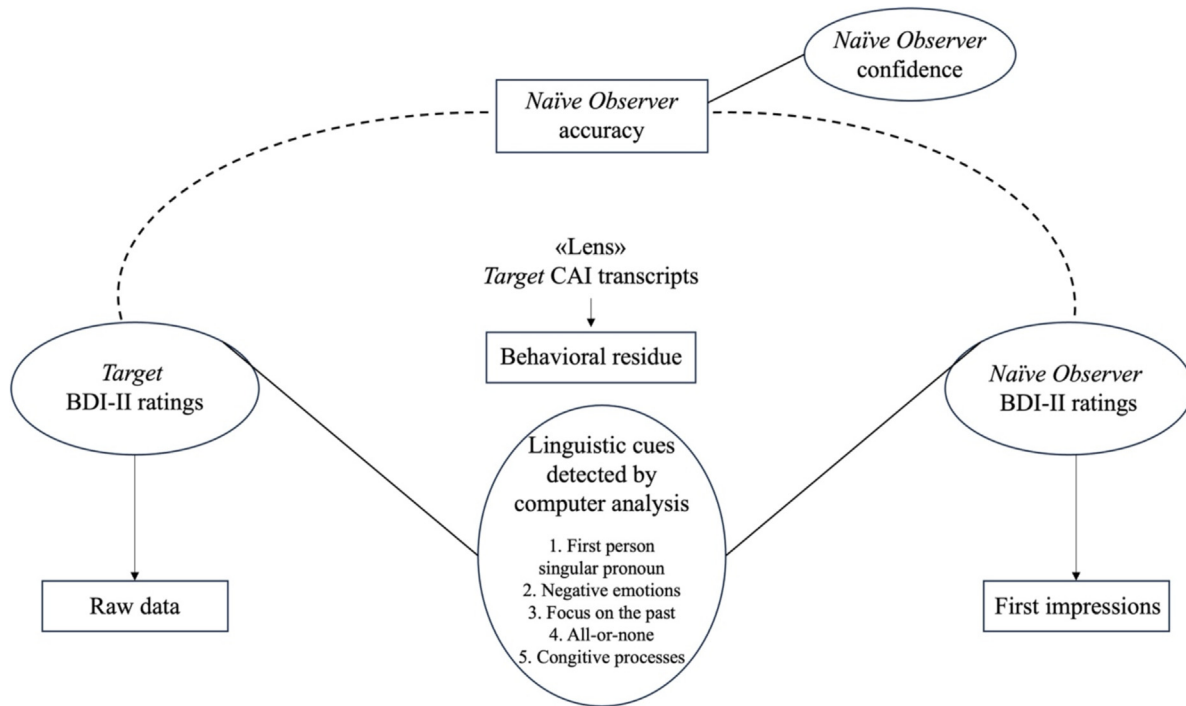


FIGURE 1 Brunswik's lens model applied to the current investigation (Gosling et al., 2002). BDI-II = Beck Depression Inventory; CAI = Child Attachment Interview.

DATA ANALYSES

To test study hypotheses, we conducted analyses using SPSS Statistics for MAC, Version 28 (IBM Corp., Armonk, NY, USA), and jamovi software (deidentified data, analysis, and research materials are available under request). Using independent samples *t*-tests, analysis of variance (ANOVAs), and correlations, the associations between research variables and demographic factors were evaluated. We present the results separately by round for Aims 1, 2 and 3. First we conducted tests of skewness and kurtosis; all data were normally distributed.

To test Aim 1, we calculated reliability assessments (Intraclass Correlation Coefficient), i.e., ICC (3,k) (Shrout & Fleiss, 1979), for the BDI-II that the three observers rated (H1). A 95% confidence interval of ICC (3,k) estimate at least between .50 and .75 is indicative of moderate interrater reliability (Koo & Li, 2016) and would support our hypotheses.

To test Aim 2, we computed mean scores of the observers' ratings. Using those mean scores, we performed Pearson's correlation analysis to assess whether observers' ratings of targets' BDI-II were associated with targets' self-ratings (H2). Effect sizes greater than $r = .40$ (Watson et al., 2000) were considered supportive of this hypothesis.

To test Aim 3, we conducted a hierarchical linear regression for both Round 1 and 2. Covariates were included in the first step, selected linguistic word categories (i.e., first-person singular pronouns, negative emotions, focus on the past, all-or-none and cognitive processes) in the second step, and NOs' ratings in the third step (H3).

To test Aim 4, repeated measures ANOVAs were performed to evaluate whether there was a difference between the three observers' BDI-II ratings for each target (within subject factor) in the HA sample compared to the CA sample (between subject factor) (H4a). To test H4b, we conducted a linear regression model using the PROCESS macro (Model 1) for SPSS (Hayes, 2018) with 5000 bias-corrected bootstrapped samples, to evaluate whether the clinical group moderated the association between targets' ratings (independent variables) and observers' ratings (dependent variable).

To test Aim 5, we conducted a regression analysis with observers' confidence (independent variable) and targets' self-rating for depression (dependent variable) to assess their association (H5a). For H5b we conducted a regression to evaluate whether the group predicted observers' confidence, controlling for the length of hospitalization.

Results

PRELIMINARY ANALYSES

In comparing groups (HA and CA), our results revealed that HA were significantly more likely to be White than CA, $\chi^2 = 94.70$, $p < .001$, and parents' education differed across groups, $\chi^2 = 21.70$, $p = .001$, with a higher number of HA parents falling into the "high school diploma" and "PhD/Doctoral" categories compare to CA parents. There were no significant differences between the two groups in gender, $\chi^2 = 1.20$, $p = 0.27$, or age, $t(190) = -1.17$, $p > .05$. Consistent with expectations, an independent samples Welch t -test revealed that HA reported significantly greater depressive symptoms than CA, $t(152) = -8.82$, $p < .001$ (M_{tot} HA = 23.70, $SD = 14.20$, M_{tot} CA = 8.87, $SD = 8.31$). Fifty-six percent of HA scores fell above the clinical cut-off for depressive disorders in adolescence on the BDI-II (raw score ≥ 23 for moderate depression symptoms; Dolle et al., 2012), while 5.1% of CA scored above this cut-off.

In order to explore potential confounding variables, we examined associations among demographic variables and key study variables. Correlations between targets' age and BDI-II total score did not show any significant relationship ($p > .05$). Regarding race and ethnicity, in the HA sample there were no differences across these groups with BDI-II total score ($p > .05$), while in the CA sample, White adolescents scored higher than Non-White adolescents ($F = 8.18$, $p = 0.01$; $M_{White} = 9.36$, $SD = 8.50$; $M_{Non-White} = 4.10$, $SD = 5.07$). There was also a significant interaction between gender and group in BDI-II scores ($F = 5.39$, $p = 0.02$). In the CA sample females and males did not differ ($M_{Females} = 9.43$, $SD = 8.92$; $M_{Males} = 7.20$, $SD = 6.02$; $p = 0.83$), while in the HA sample females scored significantly higher than males ($M_{Females} = 27.10$, $SD = 13.80$; $M_{Males} = 16.30$, $SD = 12.90$; $p < .001$). Parent education was not associated with BDI-II score ($p > .05$). Independent samples t -tests revealed that HA adolescents above the BDI-II clinical cut-off were hospitalized longer than those below the cut-off in both Round 1 ($M_{Below\ cut-off} = 27.80$, $SD = 7.40$; $M_{Above\ cut-off} = 32.30$, $SD = 10.60$; $p = 0.02$) and Round 2 ($M_{Below\ cut-off} = 27.40$, $SD = 7.61$; $M_{Above\ cut-off} = 32.20$, $SD = 10.40$; $p = 0.01$). As a result of these findings, gender, ethnicity and length of hospitalization were included as covariates in subsequent analysis with BDI-II.

As part of our data exploration, we also examined bivariate correlations between key study variables (see Table 2). In order to explore differences

in the language use of HA and CA on the CAIs, we created three custom LIWC dictionaries consisting of words pertaining to mental health, hospitalization, and depression (see Supplemental documents, Section 2, for more information). Results showed that the CAIs of HA had significantly higher scores than CA on all three categories: mental health, $p < .001$, Cohen's $d = 0.33$, hospitalization, $p = .05$, Cohen's $d = 0.16$, and depression, $p < .001$, Cohen's $d = 0.29$, all three with small effect sizes.

HYPOTHESIS TESTING

Aim 1-H1: Are the observers internally consistent in their evaluations of adolescents' symptomatology after reading attachment transcripts?

In both Round 1 and Round 2, ICC testing for the BDI-II across three raters revealed moderate reliability (Round 1 $\alpha = 0.74$; Round 2 $\alpha = 0.68$). These results were consistent with expectations, (value $> .50$).

Aim 2-H2: Do observers' ratings of targets' depression converge with targets' rating of their depression?

In Round 1, observers' ratings were significantly positively associated with targets' ratings of their depressive symptoms ($r = 0.53$, $p < .001$), reaching the cut-off we expected ($r = 0.40$). Likewise, a regression analysis revealed that observers' ratings of depression were significant predictors of targets' ratings after controlling for gender, ethnicity and length of hospitalization, $\Delta R^2 = 0.06$, $\beta = 0.36$, $p < .001$. In Round 2, observers' ratings were significantly positively associated with targets' ratings of their depressive symptoms ($r = 0.58$, $p < .001$). Once again, the significance held in a regression controlling for targets' gender, ethnicity, and length of stay, $\Delta R^2 = 0.10$, $\beta = 0.28$, $p < .001$.

Aim 3-H3: Do NOs' ratings provide an incremental contribution in predicting targets' ratings of depression compared to the psycholinguistic cues alone?

In Round 1 and 2, correlation analysis revealed that only cognitive processes ($r = -.20$, $p < .01$) and focus on the past ($r = -.36$, $p < .01$) categories were significantly associated with targets' ratings. Thus, we conducted a hierarchical linear regression model. In Round 1, after including targets' gender, ethnicity, and CAs' length of hospitalization in the initial step, $R^2 = 0.65$ $F(3, 181) = 44.50$ $p < .001$, cognitive processes, focus on the past words were added in the second step, $\Delta R^2 = 0.01$, $p = .14$. In the last step NOs ratings were included, $\Delta R^2 = 0.05$, $p < .001$. Overall, the

Table 2

Zero-Order Pearson's Correlations and Descriptive Statistics for Key Study Variables in Both HA and CA Groups, $N = 195$ (Round 1), $N = 191$ (Round 2)

Variables	1	2	3	4	5	6	7	8	9
Round 1									
Mean (SD)	15.30 (1.26)	14.70 (12.10)		16.10 (13.80)					
1. Targets' age	—	—		—					
2. NOs' BDI-II rating	.130	—		—					
4. Targets' BDI-II	.010	.525***		—					
Round 2									
Mean (SD)	15.30 (1.26)	14.20 (11.60)	4.01 (0.53)	17.40 (13.80)	17.30 (2.75)	1.52 (0.62)	6.04 (1.97)	1.65 (0.70)	13.60 (2.99)
1. Targets' age	—	—		—					
2. NOs' BDI-II rating	-.028	—		—					
3. NOs' BDI-II confidence	-.063	-.207**	—	—					
4. Targets' BDI-II	.055	.326***	-.084	—					
5. 1 st p. sg. pronouns	.071	.113	-.059	.009	—				
6. Negative emotions	.011	-.029	.025	-.109	.058	—			
7. Focus on the past	.032	-.307***	.021	-.209**	.186*	.436***	—		
8. All-or-none	.015	.081	-.037	.024	.134	.276***	.121	—	
9. Cognitive processes	-.058	-.236**	.149*	-.236**	.271***	.356***	.344***	.111	—

Note. BDI-II = Beck Depression Inventory-II Edition; NOs = Naïve Observers; 1st p. sg. pronouns = First person singular pronouns.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

complete model explained about 47.5% of the variability in the targets' ratings revealing that: gender was a significant predictor of targets' self-report depression symptomatology, $\beta = -5.63$, $SE = 1.72$, $p = .001$, and length of hospitalization, $\beta = 0.53$, $SE = 0.06$, $p < .001$, while ethnicity was not significant, $\beta = -1.32$, $SE = 2.09$, $p > .05$. When included in the second step, focus on the past was significantly associated, $\beta = -0.88$, $SE = 0.44$, $p = .047$, while cognitive processes were not, $\beta = 0.12$, $SE = 0.28$, $p > .05$. When NOs' ratings were added in the final step, they were significantly associated with targets' ratings, $\beta = 0.35$, $SE = 0.80$, $p < .001$, while focus on the past ($\beta = -0.66$, $SE = 0.42$, $p > .05$) lost its significance. In Round 2, after including targets' gender, ethnicity, and CAs' length of hospitalization in the initial step, $R^2 = 0.64$, $F(3, 175) = 41.10$, $p < .001$, cognitive processes and focus on the past words were added in the second step, $\Delta R^2 = 0.01$, $p = .33$. In the last step NOs' ratings were included, $\Delta R^2 = 0.08$, $p < .001$. Overall, the complete model explained about 49.7% of the variability in the targets' ratings, revealing that: gender was a significant predictor of targets' self-report depression symptomatology, $\beta = -5.86$, $SE = 1.63$, $p < .001$, and length of hospitalization, $\beta = 0.41$, $SE = 0.07$, $p < .001$, while ethnicity was not significant, $\beta = -2.96$, $SE = 2.02$, $p > .05$. When included in the second step, neither focus on the past,

$\beta = -0.68$, $SE = 0.46$, $p = .14$, nor cognitive processes, $\beta = 0.08$, $SE = 0.28$, $p = .78$, were associated with targets' ratings. When NOs' ratings were added in the final step, they were significantly associated with targets' ratings, $\beta = 0.48$, $SE = 0.09$, $p < .001$, while cognitive processes and focus on the past remained nonsignificant. Analyses in both Round 1 and 2 suggest the incremental validity of NOs' ratings above and beyond the automatic psycholinguistic cues.

Aim 4-H4a: Do observers' ratings for depression differ between HA and CA groups?

In Round 1, a repeated measure ANOVA revealed that there was a difference in observers' detection of depression between HA and CA, $F(1, 193) = 100.64$, $p < .001$, with 34% of the variance in observers' ratings explained by group membership. Observers detected higher depressive symptoms in HA compared to CA ($M_{HA} = 21.90$, $SD_{HA} = 11.70$; $M_{CA} = 7.73$, $SD_{CA} = 7.49$). In Round 2, a repeated measure ANOVA revealed that there was a difference between observers' rating for depression between HA and CA, $F(1, 196) = 92.64$, $p < .001$, with 32% of the variance in observers' ratings explained by group membership. Observers detected higher depressive symptoms in HA compared to CA in Round 2 as well ($M_{HA} = 18.20$, $SD_{HA} = 8.88$; $M_{CA} = 7.42$, $SD_{CA} = 6.95$). Findings confirmed our hypothesis.

Aim 4–H4b: Does the group moderate the association between observers' ratings and targets' rating?

We conducted two moderated regression analyses. In Round 1, controlling for covariates, the association between observers' and targets' ratings for depression was not moderated by group, $\Delta R^2 = 0.00$, $\beta = 0.01$, $p = 0.94$. In Round 2, controlling for covariates, the association between observers' and targets' ratings for depression was moderated by group, $\Delta R^2 = 0.17$, $\beta = -0.23$, $p = 0.02$. This association was positive and significant in both groups but was stronger in CA than in HA ($b_{CA} = 0.43$, $b_{HA} = 0.20$). While in Round 1 associations between observers' ratings and targets' self-reports did not differ by sample, in Round 2 observers' ratings and targets' self-reports were more strongly associated in CA than in HA.

Aim 5–H5a: Is observers' confidence in their perceptions of targets' depression positively associated with targets' own reports of their symptomatology?

After including targets' gender, ethnicity and HAs' length of hospitalization in the initial step, $R^2 = 0.30$, $F(3, 169) = 24.1$, $p < .001$, observers' confidence was added in the next step, $\Delta R^2 = 0.01$, $p = .51$. The model explained 32% of the variability: targets' gender was a significant predictor of targets' self-report depression symptomatology, $\beta = -5.29$, $SE = 1.91$, $p = .01$, as well as ethnicity, $\beta = 4.43$, $SE = 2.05$, $p = .03$, and length of hospitalization, $\beta = 0.34$, $SE = 0.06$, $p < .01$, while observers' confidence was not a significant predictor, $\beta = -1.07$, $SE = 1.61$, $p > .05$.

Aim 5–H5b: Is group membership associated with observers' confidence in their perceptions of targets' depression?

After including length of hospitalization in the initial step, $R^2 = 0.02$, $F(2, 186) = 4.50$, $p = .035$, the group was added in the second step, $\Delta R^2 = 7.12$, $p = .71$. The model explained 1% of the variability: neither length of hospitalization, $\beta = -0.003$, $SE = 0.005$, $p > .05$, nor group, $\beta = -0.06$, $SE = 0.15$, $p > .05$, were significant predictors of NOs' confidence in their ratings of targets' depression.

Discussion

Observers, such as peers or teachers, have the potential to offer a unique window into the detection of depressive symptoms in adolescence, both in clinical and community settings, relying solely on behavioral residue like linguistic and verbal cues. Their perceptions may coincide with those of the adolescent, or they may offer nonoverlap-

ping information. Observers may be able to notice subtle changes in language, behavior, mood, or social interactions that the youth themselves may not recognize or report. Able to capture day-to-day variations, observers are often the first to notice early-warning signs of depression, such as withdrawal from social activities, changes in academic performance, or alterations in verbal expression. These early signs might not yet meet clinical thresholds but can indicate emerging issues.

To explore the role of observers in detecting adolescents' depressive symptoms, we examined whether observers were internally consistent in their evaluations after reading attachment transcripts (Aim 1). Based on prior research on the PP paradigm, we anticipated that observers would be moderately consistent in their ratings, which was supported by our findings. Then we examined the convergence between observers' and targets' ratings (Aim 2). Observers' ratings were significantly and positively correlated with targets' ratings of depressive symptoms. Our findings revealed remarkable congruence between observer and target ratings. Notably the congruence was significant but not absolute, as anticipated by the multi-informant perspective (Kraemer et al., 2003) and consistent with prior work on the detection of internalizing symptoms (for a review, De Los Reyes et al., 2015). However, this study takes a step further, through the use of a group of unfamiliar observers who were completely unaware of any other cues related to the target they were evaluating. Notably, these observers were not trained in detecting mental health or depressive features, yet they were able to moderately accurately detect signs of distress/depression in adolescents with the same statistical accuracy as familiar informants from previous research, who had many more elements on which to base their reports. In addition, observers were able to detect depressive symptoms with moderate accuracy not only in the HA, with potentially higher symptom intensity, but also in the CA.

In Aim 3, we found that NOs ratings predicted targets' ratings above and beyond automatic linguistic analysis alone. This may suggest that observers provide additional insights beyond what is obtained through computerized analysis. Interestingly, in Round 1, our linguistic analysis revealed that higher levels of the "focus on the past" word category were associated with lower levels of BDI-II rated by targets, which differs from prior findings (Cohn et al., 2004; Lyons et al., 2018; Rude et al., 2004). This pattern of findings raises interesting questions regarding what naïve observers

pick up on beyond specific word frequencies. Further research delving into these associations is indicated.

We later examined whether evaluations made by the observers differed for HA compared to the CA (Aim 4a) and if these evaluations showed a higher degree of accuracy in the HA (Aim 4b). Indeed, observers achieved greater symptom detection for depression in HA. Youth with depressive features may use a distinctive narrative-verbal mode (e.g., [Suherman et al., 2018](#)), which, more conspicuously, could also enable an untrained observer to detect potential risk for depression. Based on preliminary analyses using three linguistic categories (mental health, hospitalization and depression-related words), it is possible higher rates of these words may have been the behavioral residue that enabled observers to detect higher depressive symptoms in HA relative to CA.

Our second assumption, that the association between targets' and observers' ratings was moderated by the group, was based on the idea that hospital settings would exert a "priming effect" on targets, leading them to discuss depressive symptoms more openly and provide greater and/or more relevant linguistic cues for observers to evaluate. Note that the linguistic analysis we conducted in our preliminary analysis supported the conclusion that HA used more depression-related, mental health, and hospitalization-related words than CA. Despite this, results did not confirm this second assumption. Counterintuitively, in Round 2, we found greater observer accuracy in rating CA. The challenge may not lie so much in determining whether the adolescent is symptomatic or not (indeed, our data confirm that this was detected), but rather in defining the severity of symptoms. The diagnostic heterogeneity of the HA group might have contributed to a weakening of the association between observers' evaluations and the hospitalized targets' evaluations. From a PP perspective ([Gosling et al., 2002](#)), our results could signify that, in hospitalized samples, observers are capable of detecting different information from the behavioral residues that do not always correspond to what the patient reports. It may help to clarify the ultimate assessment goal: we should not be striving for perfect agreement but even forcing disagreement in the process. This is consistent with the multi-informant perspective within the mental health literature ([Kraemer et al., 2003](#)), which holds that discrepancies in different raters' assessments of people's mental health symptoms provide meaningful information. Each observer brings their own unique perspective that allows them to perceive information about the target that

is relevant and important. This results in each observer contributing unique information, triangulating on perceptions that are different but form a complementary view on the person's trait, detecting different facets of youth mental health. Even among observers, there may be differences based on the information each one considers salient, and this could be the reason why observers' evaluations were similar but not identical (see results in Aim 1). Further, in our study, HA adolescents self-reported a higher level of distress compared to that reported by observers. This is not surprising, as it is not uncommon for inpatients, especially those with personality disorders, to "inflate" their symptoms, perhaps as a plea for help and attention to their suffering ([Boberg et al., 2023](#)). Despite this, 5.1% of CA in our sample exceeded the BDI-II cutoff for mild depression, confirming our assumption that depressive symptoms were also present in the CA.

Finally, we were interested in the observers' subjective perspective of their confidence and evaluation (Aim 5). Analyses revealed that observers' confidence in their evaluations was not associated with either targets' ratings (H5a) or group membership (H5b). Future studies should examine other predictors of confidence, as well as its downstream behavioral consequences.

IMPLICATIONS AND FUTURE DIRECTIONS

Our work contributes to the body of literature demonstrating how thin slices of information—i.e., verbal transcripts with emotional content—can reveal information detectable by others. Specifically, emotional content of adolescents' verbal cues is detectable in many adolescent conversations and interactions ([Flannery et al., 1994](#)). In this study, we assumed that the questions of the CAI might create an ideal context in which to observe adolescents' emotional states. This study's findings may partially be attributed to the salience of cues taken from the adolescents' interviews. The CAI elicits relational memories involving primary figures and stimulates episodes related to moments of discouragement, loss, or separation (e.g., "*Can you tell me about a time when you were really upset and wanted help?*"). These questions, though not specifically related to mental health episodes, could have elicited dysphoric content in our sample, which might have facilitated the detection process. From the observer's perspective, reading transcripts of the CAI is not as ecologically valid as observing a spontaneously occurring conversation involving the target. In other words, the findings may be partially constrained by the interview on which the observers relied for their perceptions.

To overcome this potential limitation, future studies can explore the ecological validity of observer detection by using more naturalistic contexts. In a real-world scenario, observers may rely on natural language use in different contexts, such as conversations, social media posts, or written communication, as a substitute for the transcripts used in our study. This approach would allow them to identify symptom intensity in a more realistic context through everyday language patterns and/or content before they reach a cut-off of severity. In addition, integrating nonverbal cues, such as facial expressions and body language, could provide a more comprehensive understanding of how depressive symptoms are perceived by others. This approach may help determine whether integrating multiple types of behavioral cues in different contexts improves the accuracy of detecting mental health problems. For instance, it could explore social environments where adult informants are generally absent (e.g., peer interactions) or utilize video and audio recordings to analyze the impact of nonverbal cues on detection of internalizing symptoms.

Given the current study's use of attachment-based narratives to detect linguistic cues, it will also be interesting to compare the discussion of positive topics versus negative experiences (e.g., *"tell me about your best day"* vs *"your worst nightmare"*) to study what can be perceived by others based on preselected narrative contexts and other nonattachment contexts. This kind of research could also be extended to other internalizing symptoms, like anxiety, seeking to understand whether and from which type of behavioral residue they can be inferred by an observer and whether observers can detect internalizing symptoms (depression, anxiety) that do not co-occur with other symptoms.

If these findings are replicated and extended through future research, they could have practical implications for prevention. There is a precedent for the notion that peers could assist in helping to detect depressive symptoms and in directing teens to seek help (see also Cannon et al., 2020 on social anxiety). If further refined and better understood, it is possible that observers' input could improve the prevention process, enhancing confidence that outreach to mental health professionals is a useful next step. Previous research documented convergence from different informants (i.e., parents, adolescents, clinicians) with ratings provided by unfamiliar independent observers about externalizing and internalizing problems (Glenn et al., 2019; Rezeppa et al., 2021; Makol et al., 2020). A multi-informant approach has

practical significance, advocating for tailored assessment methods wherein making decisions without thoroughly integrating reports from multiple informants can lead to suboptimal treatments (De Los Reyes et al., 2022). In future work, it would be interesting to compare observers' accuracy based on reading transcripts to clinicians' accuracy in rating targets' depressive symptoms after conducting a clinical interview with them.

This study suggested that observers were able to distinguish HAs from CAs with respect to depressive symptoms. We know from previous studies that adolescents without a mental health diagnosis may exhibit depressive symptoms (e.g., Shorey et al., 2022), which might be more difficult to detect than those expressed by a hospitalized sample due to less degree of intensity. Future research could explore whether observers are also able to discriminate between CA and an outpatient sample of adolescents with depression through the emotional content of adolescents' verbal cues, that are detectable in many adolescent conversations and interactions (Flannery et al., 1994). There have been efforts in recent years to emphasize the benefits of peer supports (e.g., peer educators, peer mentors) among adolescents and college students as early points of detection or intervention, especially among clinical populations (for reviews, see Merianos et al., 2016; Murphy et al., 2024)—these individuals are relatable, have shared experience, and have easy access to at-risk students. Perhaps learnings from this study and others that follow it can be used to enhance these types of programs—place students in the right situations where they disclose relevant content and peers (observers) may be able to detect risk.

Moreover, our findings indicate that in Round 2, observers were more accurate in detecting CA's depressive symptoms. This suggests that our findings (that untrained observers are accurate in identifying depressive symptoms from transcripts) may have more applicability to community settings. This is promising given the need for effective prevention strategies and increasing reliance on peers as mental health interveners and detectors (e.g., Chahl et al., 2024; Roach, 2018).

Although the focus of this work involved human observers, another potential direction will be to further explore whether AI or linguistic software can identify specific linguistic categories most strongly associated with the depressive symptoms identified in the targets and confirm that the observer has stronger explanatory potential compared to AI. On this point, observers provide a more ecologically valid yet resource-intensive assessment option, but this could change over time as the

acceptability of technology increases. Both raise ethical concerns in the real-world scenario—asking peers to report on other peers can impinge on privacy, as can using technology to record and analyze speech. It is clear that future research in this area is needed and offers exciting new opportunities for the early detection of mental health concerns.

Last, in this study, we did not consider observers' characteristics, but previous studies (e.g., Borelli et al., 2019) assessed how the observers' traits, such as their own attachment, can influence their evaluations. Expanding on the ways in which we form initial impressions of others based on individual characteristics will enable us to better comprehend how psychopathological symptoms are perceptible to others, to what extent, and for whom.

LIMITATIONS

The contributions of this study should be considered in light of its limitations. First, this study set out to explore the role of naïve observers in detecting adolescents' depression from subtle linguistic cues. While we provide preliminary support for this effect, the clinical applicability of these findings remains unknown. Moreover, the study was cross-sectional, partially exploratory, and not pre-registered. The generalizability of findings is reduced due to the racial/ethnic homogeneity and limited demographic information of the sample overall as well as the clinical acuity of the HA sample. Observers rated transcripts rather than conversations they overheard, so future work should examine whether accuracy is as great when observers are presented with oral speech samples. In addition, the use of two such different samples (HA and CA) can be seen as a strength (it gave observers the opportunity to accurately detect depressive symptoms) but also a limitation (findings may be accounted for by severity of symptoms), underscoring the need for follow-up work comparing CA to outpatient samples. The comorbidities also introduce the possibility that the presence of other symptoms could have influenced the observers' ratings. Moreover, we did not have measures to detect other potentially comorbid symptoms, such as anxiety, that could have been used as a point of comparison. Finally, the linguistic analysis we performed only examined a few language categories that prior research has connected to depressive symptoms. Other linguistic areas could be investigated, such as mental health words (e.g., depressed, suicide, trauma) or swear words.

CONCLUSION

Unfamiliar and untrained observers are able to reliably and accurately detect depressive symptoms in both HA and CA using transcribed linguistic cues of emotionally evocative interviews. The human observers' ratings predict targets' ratings after controlling for the information provided by automatic linguistic analysis. These findings lay the groundwork for future studies poised to delve more deeply into questions related to the prevention of mental health problems in adolescence.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.beth.2024.12.002>.

References

- Achenbach, T. M., Dumenci, L., & Rescorla, L. (2002). Ten-year comparisons of problems and competencies for national samples of youth: self, parent and teacher reports. *Journal of Emotional and Behavioral Disorders*, 10(4), 194–203. <https://doi.org/10.1177/10634266020100040101>.
- 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. <https://doi.org/10.1177/2167702617747074>.
- Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013). Detecting depression: a comparison between spontaneous and read speech. In *2013 IEEE international conference on acoustics, speech and signal processing, Vancouver, BC, Canada* (pp. 7547–7551). <https://doi.org/10.1109/ICASSP.2013.6639130>.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation. https://doi.org/10.1007/978-1-4419-1005-9_441.
- Bernieri, F. J., Zuckerman, M., Koestner, R., & Rosenthal, R. (1994). Measuring person perception accuracy: another look at self-other agreement. *Personality and Social Psychology Bulletin*, 20(4), 367–378. <https://doi.org/10.1177/0146167294204004>.
- Biesanz, J. C., & Wallace, J. D. (2020). Accuracy in person perception. In *The Cambridge handbook of personality psychology*, pp. 44–55. Cambridge University Press. <https://doi.org/10.1017/9781108264822.006>.
- Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, 92(1), 119–135. <https://doi.org/10.1037/0022-3514.92.1.119>.
- Boberg, M., Jeppesen, U., Arnfred, S., & Nordgaard, J. (2023). Do we know the mind of others? Suspicion of malingering in emergency psychiatry. *Nordic Journal of Psychiatry*, 77(3), 234–239. <https://doi.org/10.1080/08039488.2022.2083676>.
- Borelli, J. L., David, D. H., Crowley, M. J., & Mayes, L. C. (2010). Links between disorganized attachment classification and clinical symptoms in school-aged children. *Journal of Child and Family Studies*, 19, 243–256. <https://doi.org/10.1007/s10826-009-9292-8>.

- Borelli, J. L., Pedroza, M., Gaskin, G. E., Smiley, P. A., Kernick, C. A., Brachman, A., & Mehl, M. R. (2018). School-aged children's cognitive interdependence as a prospective link between their depressive symptoms and physiological stress reactivity. *Journal of Social and Clinical Psychology, 37*(5), 325–355. <https://doi.org/10.1521/jscp.2018.37.5.325>.
- Borelli, J. L., Peng, X., Hong, K., Froidevaux, N. M., & Sbarra, D. A. (2019). Accuracy and confidence in perceptions of targets' attachment to former partners: do judges vary as a function of individual differences in attachment orientation? *Journal of Research in Personality, 79*, 1–12. <https://doi.org/10.1016/j.jrp.2019.01.001>.
- Cannon, C. J., Makol, B. A., Keeley, L. M., Qasmieh, N., Okuno, H., Racz, S. J., & De Los Reyes, A. (2020). A paradigm for understanding adolescent social anxiety with unfamiliar peers: conceptual foundations and directions for future research. *Clinical Child and Family Psychology Review, 23*(3), 338–364. <https://doi.org/10.1007/s10567-020-00314-4>.
- Chahl, A., Bai, S., Rulison, K. L., & Fosco, G. M. (2024). Examining the association between school connectedness and use of self-regulation strategies in middle childhood. *Applied Developmental Science, 1–10*. <https://doi.org/10.1080/10888691.2024.2305343>.
- Charamut, N. R., Racz, S. J., Wang, M., & De Los Reyes, A. (2022). Integrating multi-informant reports of youth mental health: a construct validation test of Kraemer and colleagues' (2003) Satellite Model. *Frontiers in Psychology, 13*, 911629. <https://doi.org/10.3389/fpsyg.2022.911629>.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*(10), 687–693. <https://doi.org/10.1111/j.0956-7976.2004.00741.x>.
- Costello, J. E., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry, 47*(12), 1263–1271. <https://doi.org/10.1111/j.1469-7610.2006.01682.x>.
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858–900. <https://doi.org/10.1037/a0038498>.
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123–149. <https://doi.org/10.1146/annurev-clinpsy-050212-185617>.
- De Los Reyes, A., Talbott, E., Power, T. J., Michel, J. J., Cook, C. R., Racz, S. J., & Fitzpatrick, O. (2022). The needs-to-goals gap: how informant discrepancies in youth mental health assessments impact service delivery. *Clinical Psychology Review, 92*, 102114. <https://doi.org/10.1016/j.cpr.2021.102114>.
- De Los Reyes, A., Wang, M., Lerner, M. D., Makol, B. A., Fitzpatrick, O., & Weisz, J. R. (2023). The operations triad model and youth mental health assessments: catalyzing a paradigm shift in measurement validation. *Journal of Clinical Child and Adolescent Psychology, 52*(1), 19–54. <https://doi.org/10.1080/15374416.2022.2111684>.
- Demyttenaere, K., & Heirman, E. (2020). The blurred line between anxiety and depression: hesitations on comorbidity, thresholds and hierarchy. *International Review of Psychiatry, 32*(5–6), 455–465. <https://doi.org/10.1080/09540261.2020.1764509>.
- Deros, D. E., Racz, S. J., Lipton, M. F., Augenstein, T. M., Karp, J. N., Keeley, L. M., Qasmieh, N., Grewe, B. I., Aldao, A., & De Los Reyes, A. (2018). Multi-informant assessments of adolescent social anxiety: adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy, 49*(1), 84–98. <https://doi.org/10.1016/j.beth.2017.05.001>.
- Dolle, K., Schulte-Körne, G., O'Leary, A. M., von Hofacker, N., Izat, Y., & Allgaier, A. K. (2012). The Beck Depression Inventory-II in adolescent mental health patients: cut-off scores for detecting depression and rating severity. *Psychiatry Research, 200*(2–3), 843–848. <https://doi.org/10.1016/j.psychres.2012.05.011>.
- Flannery, D. J., Torquati, J. C., & Lindemeier, L. (1994). The method and meaning of emotional expression and experience during adolescence. *Journal of Adolescent Research, 9*(1), 8–27. <https://doi.org/10.1177/074355489491002>.
- Fleming, T., Merritt, R., McKeown, M., & Sheridan, J. (2019). Exploring the relationship between child mental health and behavioral problems using self-report and caregiver data. *Journal of Child and Adolescent Social Work, 31*(1), 55–68. <https://doi.org/10.1007/s10560-018-0581-5>.
- Follet, L., Okuno, H., & De Los Reyes, A. (2023). Assessing peer-related impairments linked to adolescent social anxiety: strategic selection of informants optimizes prediction of clinically relevant domains. *Behavior Therapy, 54*(1), 29–42. <https://doi.org/10.1016/j.beth.2022.06.010>.
- Forgas, J. P., & Bower, G. H. (1987). Mood effects on person-perception judgments. *Journal of Personality and Social Psychology, 53*(1), 53–60. <https://doi.org/10.1037/0022-3514.53.1.53>.
- Fry, M., MacGregor, C., Hyland, S., Payne, B., & Chenoweth, L. (2015). Emergency nurses' perceptions of the role of confidence, self-efficacy and reflexivity in managing the cognitively impaired older person in pain. *Journal of Clinical Nursing, 24*(11–12), 1622–1629. <https://doi.org/10.1111/jocn.12763>.
- Funder, D. C. (2018). *Accuracy in personality judgment a (very) long view. Paper presented at the Wake Forest College Colloquium series, Winston-Salem, NC.*
- Funkhouser, C. J., Trivedi, E., Li, L. Y., Helgren, F., Zhang, E., Sritharan, A., Cherner, R. A., Pagliaccio, D., Durham, K., Kyler, M., Tse, T. C., Buchanan, S. N., Allen, N. B., Shankman, S. A., & Auerbach, R. P. (2023). Detecting adolescent depression through passive monitoring of linguistic markers in smartphone communication. *Journal of Child Psychology and Psychiatry, 65*(7), 932–941. <https://doi.org/10.1111/jcpp.13931>.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (Vol. 2 (4th ed.), pp. 89–150). McGraw-Hill.
- Glenn, L. E., Keeley, L. M., Szollos, S., Okuno, H., Wang, X., Rausch, E., Deros, D. E., Karp, J. N., Qasmieh, N., Makol, B. A., Augenstein, T. M., Lipton, M. F., Racz, S. J., Scharfstein, L., Beidel, D. C., & De Los Reyes, A. (2019). Trained observers' ratings of adolescents' social anxiety and social skills within controlled, cross-contextual social interactions with unfamiliar peer confederates. *Journal of Psychopathology and Behavioral Assessment, 41*(1), 1–15. <https://doi.org/10.1007/s10862-018-9676-4>.
- Goldberg, S. (2014). *Attachment and development* (1st ed.) Routledge. <https://doi.org/10.4324/9780203783832>.
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. *Journal of Personality and Social*

- Psychology*, 82(3), 379–398. <https://doi.org/10.1037/0022-3514.82.3.379>.
- Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: quantification, inference, and interpretation. *Communication Monographs*, 85(1), 4–40. <https://doi.org/10.1080/03637751.2017.1352100>.
- Hurtig, M. C., & Pichevin, M. F. (1990). Salience of the sex category system in person perception: contextual variations. *Sex Roles*, 22(5), 369–395. <https://doi.org/10.1007/BF00288340>.
- Ibraheim, M., Kalpakci, A., & Sharp, C. (2017). The specificity of emotion dysregulation in adolescents with borderline personality disorder: comparison with psychiatric and healthy controls. *Borderline Personality Disorder and Emotion Dysregulation*, 4(1), 1–9. <https://doi.org/10.1186/s40479-017-0052-x>.
- Jamovi (2023). Version 2.3.21.0. <https://www.jamovi.org> [Computer program]
- Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: a social relations analysis. *Psychological Bulletin*, 102(3), 390–402. <https://doi.org/10.1037/0033-2909.102.3.390>.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160(9), 1566–1577. <https://doi.org/10.1176/appi.ajp.160.9.1566>.
- Kumar, G., Steer, R. A., Teitelman, K. B., & Villacis, L. (2002). Effectiveness of Beck Depression Inventory–II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment*, 9(2), 164–170. <https://doi.org/10.1177/10791102009002007>.
- Lenhart, A., Ling, R., Campbell, S., & Purcell, K. (2010). Teens and mobile phones: Text messaging explodes as teens embrace it as the centerpiece of their communication strategies with friends. *Pew Internet & American Life Project*. <https://ericed.gov/?id=ED525059>.
- Leussis, M. P., & Andersen, S. L. (2008). Is adolescence a sensitive period for depression? Behavioral and neuroanatomical findings from a social stress model. *Synapse*, 62(1), 22–30. <https://doi.org/10.1002/syn.20462>.
- Low, L. S. A., Maddage, N. C., Lech, M., Sheeber, L. B., & Allen, N. B. (2010). Detection of clinical depression in adolescents' speech during family interactions. *IEEE Transactions on Biomedical Engineering*, 58(3), 574–586. <https://doi.org/10.1109/TBME.2010.2091640>.
- Lyons, M., Aksayli, N. D., & Brewer, G. (2018). Mental distress and language use: linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87, 207–211. <https://doi.org/10.1016/j.chb.2018.05.035>.
- Lu, W. (2019). Adolescent depression: national trends, risk factors, and healthcare disparities. *American Journal of Health Behavior*, 43(1), 181–194. <https://doi.org/10.5993/AJHB.43.1.15>.
- Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: how conceptual and measurement models may address long-standing problems in clinical decision-making. *Clinical Psychological Science*, 8(6), 953–970. <https://doi.org/10.1177/2167702620924439>.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>.
- Melton, T. H., Croarkin, P. E., Strawn, J. R., & McClintock, S. M. (2016). Comorbid anxiety and depressive symptoms in children and adolescents: a systematic review and analysis. *Journal of Psychiatric Practice*, 22(2), 84–98. <https://doi.org/10.1097/PRA.0000000000000132>.
- Merianos, A. L., King, K. A., Vidourek, R. A., & Nabors, L. A. (2016). Mentoring and peer-led interventions to improve quality of life outcomes among adolescents with chronic illnesses. *Applied Research in Quality of Life*, 11(3), 1009–1023. <https://doi.org/10.1007/s11482-015-9415-x>.
- Mojtabai, R., Olfson, M., & Mechanic, D. (2002). Perceived need and help-seeking in adults with mood, anxiety, or substance use disorder. *Archives of General Psychiatry*, 59(1), 77–84. <https://doi.org/10.1001/archpsyc.59.1.77>.
- Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: the salience and significance of age in social judgments. *Advances in experimental social psychology* (Vol. 30, pp. 93–161). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60383-4](https://doi.org/10.1016/S0065-2601(08)60383-4).
- Murphy, R., Huggard, L., Fitzgerald, A., Hennessy, E., & Booth, A. (2024). A systematic scoping review of peer support interventions in integrated primary youth mental health care. *Journal of Community Psychology*, 52(1), 154–180. <https://doi.org/10.1002/jcop.23090>.
- National Health Service (2021). *Mental health of children and young people in England, 2021 – wave 2 follow up to the 2017 survey*. <https://digital.nhs.uk/data-and-information/publications/statistical/mentalhealth-of-children-and-young-people-in-england/2021-follow-up-to-the-2017-survey>.
- Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory–II in nonclinical adolescent samples. *Journal of Clinical Psychology*, 64(1), 83–102. <https://doi.org/10.1002/jclp.20433>.
- Osman, A., Kopper, B. A., Barrios, F., Gutierrez, P. M., & Bagge, C. L. (2004). Reliability and validity of the Beck depression inventory—II with adolescent psychiatric inpatients. *Psychological Assessment*, 16(2), 120–132. <https://doi.org/10.1037/1040-3590.16.2.120>.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC*. Austin, TX: LIWC.net [Computer software].
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>.
- Rausch, E., Racz, S. J., Augenstein, T. M., Keeley, L., Lipton, M. F., Szollos, S., Riffle, J., Moriarity, D., Kromash, R., & De Los Reyes, A. (2017). A multi-informant approach to measuring depressive symptoms in clinical assessments of adolescent social anxiety using the Beck Depression Inventory–II: convergent, incremental, and criterion-related validity. *Child & Youth Care Forum*, 46(5), 661–683. <https://doi.org/10.1007/s10566-017-9403-4>.
- Rezeppa, T., Okuno, H., Qasmieh, N., Racz, S. J., Borelli, J. L., & De Los Reyes, A. (2021). Unfamiliar untrained observers' ratings of adolescent safety behaviors within social interactions with unfamiliar peer confederates. *Behavior Therapy*, 52(3), 564–576. <https://doi.org/10.1016/j.beth.2020.07.006>.
- Rickelt, J., Viechtbauer, W., Lieverse, R., Overbeek, T., van Balkom, A. J., van Oppen, P., van den Heuvel, O. A.,

- Marcelis Eikelenboom, M., Tibi, L., & Schruers, K. R. (2013). The relation between depressive and obsessive-compulsive symptoms in obsessive-compulsive disorder: results from a large, naturalistic follow-up study. *Journal of Affective Disorders*, 203, 241–247. <https://doi.org/10.1016/j.jad.2016.06.009>.
- Roach, A. (2018). Supportive peer relationships and mental health in adolescence: an integrative review. *Issues in Mental Health Nursing*, 39(9), 723–737. <https://doi.org/10.1080/01612840.2018.1496498>.
- Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: the lay assessment of subclinical depression from written self-descriptions. *Journal of Personality*, 78(2), 575–598. <https://doi.org/10.1111/j.1467-6494.2010.00627.x>.
- Rohde, P. (2009). Comorbidities with adolescent depression. In S. Nolen-Hoeksema & L. M. Hilt (Eds.), *Handbook of depression in adolescents* (pp. 139–177). Routledge/Taylor & Francis Group.
- Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121–1133. <https://doi.org/10.1080/02699930441000030>.
- Russo, L. N., & Borelli, J. L. (2023). College women's perceptions of judgments on aggression and risk of intimate partner violence perpetration in potential romantic partners. *Journal of Family Violence*, 39(5), 835–848. <https://doi.org/10.1007/s10896-023-00531-9>.
- Schubert, K. O., Clark, S. R., Van, L. K., Collinson, J. L., & Baune, B. T. (2017). Depressive symptom trajectories in late adolescence and early adulthood: a systematic review. *Australian and New Zealand Journal of Psychiatry*, 51(5), 477–499. <https://doi.org/10.1177/0004867417700274>.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 28–38. <https://doi.org/10.1097/00004583-200001000-00014>.
- Shmueli-Goetz, Y., Target, M., Fonagy, P., & Datta, A. (2008). The child attachment interview: a psychometric study of reliability and discriminant validity. *Developmental Psychology*, 44(4), 939–956. <https://doi.org/10.1037/0012-1649.44.4.939>.
- Shorey, S., Ng, E. D., & Wong, C. H. (2022). Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. *British Journal of Clinical Psychology*, 61(2), 287–305. <https://doi.org/10.1111/bjc.12333>.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>.
- Smetana, J. G. (2008). “It’s 10 o’clock: Do you know where your children are?” Recent advances in understanding parental monitoring and adolescents’ information management. *Child Development Perspectives*, 2(1), 19–25. <https://doi.org/10.1111/j.1750-8606.2008.00036.x>.
- Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: how (and how well) people use others’ opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 780–805. <https://doi.org/10.1037/a0015145>.
- Stuckey, H., Hivner, E. A., Kraschnewski, J. L., Molinari, A. M., Costigan, H. J., & Sekhar, D. L. (2021). “I wouldn’t even know what to do,” adolescent and parent perspectives on identifying, understanding, and seeking help for adolescent depression. *Psychiatric Quarterly*, 92(4), 1459–1472. <https://doi.org/10.1007/s11126-021-09918-7>.
- Suherman, S., Malini, H., & Renidayati, R. (2018). Identification of early warning signs of suicidal behavior among adolescents. *Elevate*, 1(2), 145–150. <https://doi.org/10.25077/elevate.1.2.145-150.2018>.
- Thompson, E., Spirito, A., Frazier, E., Thompson, A., Hunt, J., & Wolff, J. (2020). Suicidal thoughts and behavior (STB) and psychosis-risk symptoms among psychiatrically hospitalized adolescents. *Schizophrenia Research*, 218, 240–246. <https://doi.org/10.1089/cap.2019.0160>.
- UK Office for National Statistics (2020). *Young people’s wellbeing in the UK: 2020*. <https://www.ons.gov.uk/people-populationandcommunity/wellbeing/bulletins/youngpeopleswellbeingintheuk/2020#main-points>.
- Underwood, M. K., Ehrenreich, S. E., More, D., Solis, J. S., & Brinkley, D. Y. (2015). The BlackBerry Project: the hidden world of adolescents’ text messaging and relations with internalizing symptoms. *Journal of Research on Adolescence*, 25(1), 101–117. <https://doi.org/10.1111/jora.12101>.
- Venta, A., Sharp, C., & Hart, J. (2012). The relation between anxiety disorder and experiential avoidance in inpatient adolescents. *Psychological Assessment*, 24(1), 240–248. <https://doi.org/10.1037/a0025362>.
- Venta, A., Sharp, C., & Newlin, E. (2015). A descriptive study of symptom change as a function of attachment and emotion regulation in a naturalistic adolescent inpatient setting. *European Child & Adolescent Psychiatry*, 24, 95–104. <https://doi.org/10.1007/s00787-014-0532-0>.
- Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology*, 45(4), 1028–1031. <https://doi.org/10.1016/j.jesp.2009.04.026>.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self–other agreement in personality and affectivity: the role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78(3), 546–558. <https://doi.org/10.1037/0022-3514.78.3.546>.
- Whitcomb, S. A., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9780203818596>.

RECEIVED: February 12, 2024

REVISED: November 26, 2024

ACCEPTED: December 3, 2024

AVAILABLE ONLINE: 12 DECEMBER 2024