

UTILIZATION OF LINGUISTIC MARKERS IN DIFFERENTIATION OF
INTERNALIZING DISORDERS, SUICIDALITY, AND IDENTITY DISTRESS

by

ELIZABETH J. IVIE

A DISSERTATION

Presented to the Department of Psychology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Elizabeth J. Ivie

Title: Utilization of Linguistic Markers in Differentiation of Internalizing Disorders, Suicidality, and Identity Distress.

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Psychology by:

Nicholas B. Allen	Chairperson
Jennifer Pfeifer	Core Member
Kathryn Mills	Core Member
Emily Tanner-Smith	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies
-------------------	-----------------------------------

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

© 2023 Elizabeth J. Ivie

DISSERTATION ABSTRACT

Elizabeth J. Ivie

Doctor of Philosophy

Department of Psychology

June 2023

Title: Utilization of Linguistic Markers in Differentiation of Internalizing Disorders, Suicidality, and Identity Distress

The adolescent period of development is associated with a significant increase in the occurrence of mental illness. In addition, death by suicide is one of the leading causes of death amongst adolescents. Identity formation is a key developmental task of adolescence, and successful navigation of this process is associated with greater well-being and resilience, while difficulties are associated with risk for mental health disorders and suicidality. Adolescents today spend enormous amounts of time on digital devices, which have become a new instrument by which they explore and confirm their identities and experiences. The study of natural language use is related to wide range of psychological phenomena, including psychopathology, and offers a tool by which we can begin to ask and answer these questions utilizing new tools that allow us to passively collect adolescents' language use directly from their digital devices. The current study leverages a unique clinical sample of adolescents who have been followed over six months to explore the relationship between both between and within participant measures of psychopathology, suicidal thought and behaviors, and putative linguistic markers of adolescent identity formation derived from online communications in order to further understand the association between these variables using ecologically valid measures in a

community sample of adolescents experiencing significant mental health challenges. The aims of the study were to (1) assess whether there are differences in how adolescents with psychopathology, suicidal ideation, and previous suicide attempts use language, (2) language differences associated with mental illness symptomology, (3) and language differences in hypothesized identity domains associated with mental illness symptomology communicated through social communication apps via text. Participants completed baseline measures of depression, suicidality, and anxiety symptoms. Participants downloaded the EARS tool onto their digital devices that passively collected text data sent through social communication applications. The results of this study indicated that there are natural language use differences between adolescents with psychopathology and those who experience suicidality, depression, and anxiety symptoms.

CURRICULUM VITAE

NAME OF AUTHOR: Elizabeth J. Ivie

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon

DEGREES AWARDED:

Doctor of Philosophy, Clinical Psychology, 2023 , University of Oregon

Master of Science, Clinical Psychology, 2016, University of Oregon

Bachelor of Science, Psychology, 2013, University of Oregon

AREAS OF SPECIAL INTEREST:

Clinical Psychology

Depression

Adolescence

Psycholinguistic Markers

Identity Formation

PROFESSIONAL EXPERIENCE:

Predocutorial Psychology Intern, Wasatch Behavioral Health Psychology Internship Training Program, Wasatch Behavioral Health, 2022-2023

Group Leader, Dialectical Behavior Therapy Skills Group, University of Oregon Psychology Clinic, 2019-2020

External Practicum Doctoral Student Therapist, Attention Deficit Hyperactivity and Learning Disorder Assessment, University of Oregon HEDCO Clinic, 2018-2019

Doctoral Student Therapist, Integrative Behavioral Couples Therapy, University of Oregon Psychology Clinic, 2018-2019

External Practicum Doctoral Student Therapist, Early Intervention Foster Care, Oregon Community Programs, 2017-2019

Group Leader/Co-Leader, Depression Prevention Pilot, Oregon Research Institute, 2017-2018

Group Co-Leader, Sleep Intervention Pilot, ADAPT Lab, University of Oregon, 2017

Graduate Employee, Department of Psychology, University of Oregon, 2015-2022

Doctoral Student Therapist, University of Oregon Psychology Clinic, 2016-2020

PUBLICATIONS:

- Glenn, D. E., Feldman, J. S., **Ivie, E. J.**, Shechner, T., Leibenluft, E., Pine, D. S., ... & Michalska, K. J. (2021). Social relevance modulates multivariate neural representations of threat generalization in children and adults. *Developmental Psychobiology*, *63*(7), e22185.
- Ivie, E. J.**, Pettitt, A., Moses, L. J., & Allen, N. B. (2020). A meta-analysis of the association between adolescent social media use and depressive symptoms. *Journal of affective disorders*.
- Latham, M. D., Dudgeon, P., Yap, M. B., Simmons, J. G., Byrne, M. L., Schwartz, O. S., **Ivie, E.** & Allen, N. B. (2020). Factor Structure of the Early Adolescent Temperament Questionnaire–Revised. *Assessment*, *27*(7), 1547-1561.
- Michalska, K. J., Benson, B., **Ivie, E. J.**, Sachs, J. F., Haller, S. P., Abend, R., ... & Pine, D. S. (2023). Neural responding during uncertain threat anticipation in pediatric anxiety. *International Journal of Psychophysiology*, *183*, 159-170.
- Michalska, K. J., Feldman, J. S., **Ivie, E. J.**, Shechner, T., Sequeira, S., Averbeck, B., ... & Pine, D. S. (2019). Early-childhood social reticence predicts SCR-BOLD coupling during fear extinction recall in preadolescent youth. *Developmental cognitive neuroscience*, *36*, 100605.
- Giuliani, N. R., Flournoy, J. C., **Ivie, E. J.**, Von Hippel, A., & Pfeifer, J. H. (2017). Presentation and validation of the DuckEES child and adolescent dynamic facial expressions stimulus set. *International journal of methods in psychiatric research*, *26*(1).
- Markowitz, J. E., **Ivie, E.**, Kligler, L., & Gardner, T. J. (2013). Long-range order in canary song. *PLoS computational biology*, *9*(5), e1003052.

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge and express my deepest thanks to my advisor, Nick Allen. Your inquisitive nature and passion for putting psychological science into actionable steps has been a guidepost for me. My time through graduate school has been met with unforeseen obstacles and I cannot thank you enough for sticking with me with the utmost compassion and positive regard. You are truly a remarkable human being and I feel incredibly fortunate that my journey through life has included you.

Thank you to the ADAPT lab and my fellow graduate students. The comradery around the joys and challenges of graduate work, validation and encouragement, and genuine care for one another created a supportive and encouraging environment, and I will be forever grateful. Special thanks to Monika Lind, Michelle Byrne, and Ryann Crowley for making time to help me with various data analysis and coding questions at a moments notice.

Thank you to my family and friends. Thank you for being interested in the work I do and supporting me at every step. Thank you for believing in me.

This dissertation is dedicated to the children and adolescents I have had the honor to meet in my life. My deep appreciation of what you bring to the world inspires me to try to make it a better place.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	14
Literature Review.....	15
Adolescent Identity Formation.....	17
Identity Formation and Mental Illness.....	22
Natural Language Use, Psychopathology, and Identity Formation.....	36
The Current Study.....	38
Specific Aims.....	39
II. METHOD.....	44
Participants.....	44
Procedure.....	44
Materials & Measures	45
Data Analysis.....	49
III. RESULTS.....	56
Descriptive Statistics.....	56
Aim 1.....	58
Aim 2.....	61
Aim 3.....	65
IV. DISCUSSION.....	69
Aim 1.....	69
Aim 2.....	73
Aim 3.....	75

Limitations and Strengths.....	78
Future Directions.....	81
Conclusion.....	83
APPENDIX A. FIGURES.....	86
APPENDIX B. TABLES.....	93
APPENDIX C. BECK’S SCALE FOR SUICIDE IDEATION.....	95
APPENDIX D. MOOD AND FEELINGS QUESTIONNAIRE.....	96
APPENDIX E. SCREEN FOR ANXIETY RELATED EMOTIONAL DISORDERS.....	97
REFERENCES CITED.....	98

LIST OF FIGURES

Figure	Page
1. Diagnoses by Group.....	46
2. Scatterplot Total Number of Messages.....	49
3. Histograms of all Variables Prior to Data Transformation.....	86
4. Box and Whisker Plots of all Variables Prior to Data Transformation.....	87
5. QQ Plots of all Variables Prior to Data Transformation.....	88
6. Histograms of Square Root Transformed Variables.....	89
7. Normal QQ Plots of Square Root Transformed Variables.....	90
8. Histogram, Normal PP Plot, Scatterplot Residuals Aim 2 (Hyp2a).....	90
9. Histogram, Normal PP Plot, Scatterplot Residuals Aim 2 (Hyp2b)	91
10. Histogram, Normal PP Plot, Scatterplot Residuals Aim 2 (Hyp2c)	91
11. Histogram, Normal PP Plot, Scatterplot Residuals Aim 2 (Hyp2d)	91
12. Histogram, Normal PP Plot, Scatterplot Residuals Exploratory Question 1.....	92
13. Histogram, Normal PP Plot, Scatterplot Residuals Exploratory Question 2.....	92
14. Histogram, Normal PP Plot, Scatterplot Residuals Exploratory Question 3.....	92

LIST OF TABLES

Table	Page
1. (K-S) and Shapiro-Wilk Tests Prior to Data Transformation.....	53
2. (K-S) and Shapiro-Wilk Tests After Transformation.....	95
3. Overall Means and Standard Deviations.....	56
4. Means and Standard Deviations by Group.....	57
5. Box’s M Test for Homogeneity of Variance Matrices.....	59
6. Levene’s Tests of Equality of Error Variances.....	59
7. Between Subjects Effects by Group.....	60
8. Between Subjects Effects for Depression Symptoms.....	61
9. Results of Multiple Regression for Predicting Suicide Ideation Severity.....	62
10. Results of Multiple Regression Individual Predictors for Depressive Symptom Severity.....	63
11. Results of Multiple Regression Individual Predictors for STB Severity.....	64
12. Results of Multiple Regression Individual Predictors for Anxiety Symptom Severity.....	65
13. Results of Stepwise Multiple Regression Individual Predictors for STB Symptom Severity.....	66
14. Results of Stepwise Multiple Regression Individual Predictors for Depression Symptom Severity.....	67
15. Results of Stepwise Multiple Regression Individual Predictors for Anxiety Symptom Severity.....	68

CHAPTER I

INTRODUCTION

The purpose of this dissertation was to investigate language use amongst adolescents with mental illness and adolescents at-risk for suicidality. In chapter one, I begin with a review of the relevant literature on identity formation and the relationship to mental illness. Adolescence is a period in which the rates of mental illness rise substantially, and death by suicide is one of the leading causes of death for adolescents (Ruch, Sheftall, Schlagbaum, Rausch, Campo & Bridge, 2019). The primary developmental task of adolescence is the process of identity formation and research indicates that difficulty with this process contributes to the occurrence of mental illness (Berman, 2019).

Next, I reviewed the literature on natural language use and the relationship to mental illness and suicidality. Previous work identified natural language use categories that might be particularly relevant to mental illness including use of pronouns and emotional valence of words (Lyons, Aksayli & Brewer, 2018; Rude, Gortner & Pennebaker, 2004; Stamatis et al. 2022). Research on the relationship between suicidality and natural language use identified word categories including pronoun use, emotionality, and words related to death and absolutism as particularly salient (Homan et al., 2022). Much of the literature reviewed in this chapter was conducted primarily with adult samples, highlighting a need for investigation in adolescent populations as well.

At the end of chapter one, I proposed a study that investigates the relationship between natural language use, mental illness and suicidality, and identity formation in a sample of high risk adolescents. I hypothesized that there would be group differences in language use that vary as a function of the presence or absence of suicidality. I proposed exploratory investigation of questions regarding the relationship between natural language use and identity formation.

In chapter two, I described the study's methodology, including participant selection, study design, and description of study variables. In addition, I described the assessment of whether the study variables met assumptions of normality necessary to test this study's hypotheses and action taken to address issues with non-normality. In chapter three, I discussed the results of the study's hypotheses and exploratory analyses. Finally, in chapter four, I provided a synthesis of this study's findings in relation to the relevant literature. I discussed limitations and strengths of the study's findings and provide suggestions for future research.

Literature Review

Adolescence was historically termed as a period of "storm and stress" (Hall, 1905). While more recently, researchers have adopted a more wholistic view of this period of life to include many positive changes (e.g., Rich, 2003), it is notable that this period of development sees a dramatic increase in rates of mental illness, particularly anxiety and depressive disorders (Polanczyk, Salum, Sugaya, Caye & Rohde, 2015). In addition, the rates of suicidality increase substantially during adolescence and suicide is one of the leading causes of death between the ages 10-19 (Ruch, Sheftall, Schlagbaum, Rausch, Campo & Bridge, 2019). Appropriately so, there is a large body of research on

the topic of adolescent mental illness, as researchers have investigated the factors associated with difficulties navigating this period of life.

Identity development is a particularly important field of research in regard to adolescent mental illness. This period of development includes shifts in how adolescents perceive themselves in relation to the world around them (Erikson, 1966). There is an increased focus on peer relationships, particularly with increases in intimacy, both in the development of close friendships and the burgeoning of romantic interests. In addition, adolescents begin thinking more deeply about career aspirations and develop personal belief systems in regard to political and religious ideologies (Erikson, 1966). Adolescents who struggle with identity formation are also more likely to experience a range of negative outcomes, including internalizing and externalizing problems (Berman, 2019). Previous research has also shown that having a strong sense of commitment in regard to identity factors is positively related to adolescent well-being (Meeus, 2011).

The advent of digital communication technology has created a new environment in which adolescents spend considerable amount of time. Therefore, research into how adolescents use these digital spaces, and their effects, may provide important insights into risk for mental illness and identity development. Specifically, adolescents' use of language on smart devices provides an opportunity to investigate these phenomena in a naturalistic setting. This chapter will review the research on adolescent identity formation, linkages to internalizing disorders (as well as suicidality), and natural language use. I will also propose theoretical linkages between natural language use and adolescent identity formation. This chapter will conclude with a proposal to investigate

the associations between on-device natural language use and (1) anxiety and depressive disorders, (2) suicidality, and (3) identity formation.

Adolescent Identity Formation

The research on identity development throughout adolescence has grown substantially since Erikson first laid out his seminal theory of lifespan development. The conception of identity formation has expanded from the original Eriksonian concepts of “achieved vs. not achieved,” to include additional statuses of identity formation (i.e., achievement, diffusion, moratorium, and foreclosure) and also processes by which adolescents move into and out of those statuses (i.e., in-depth exploration, commitment, reconsideration of commitment) (Crocetti, 2017; Erikson, 1968; Marcia, 1966). Identity achievement describes a state in which an adolescent has explored different opportunities and ideologies and has made personal commitments. Identity diffusion describes an adolescent who has not explored options and has little interest in making commitments. Identity moratorium describes a state in which the adolescent is actively exploring options, but has not yet made commitments. Identity foreclosure describes a state in which an adolescent has made commitments in regards to occupation or ideologies, without exploration of options. These statuses were not considered continuous stages (i.e., progressing from one status to the next to eventually achieve a fully formed identity), but dynamic processes in that adolescents may oscillate between one status or another. Crocetti’s (2017) three-factor identity model included the processes of in-depth exploration, commitment and reconsideration of commitment in regards to different statuses and domains. In-depth exploration is defined as the process in which adolescents think about the commitments they have made and actively search for new information

about those commitments (Crocetti, 2017). Commitment is the state in which adolescents act consistently in various domains and feel confident about their choices.

Reconsideration of commitment is the stage in which adolescents have made commitments, but are actively thinking about other potentially "more appealing" options. They describe a cycle in which adolescents oscillate between commitment and reconsideration of commitments called the "identity formation cycle." The oscillation between commitment and in-depth exploration is the "identity maintenance cycle." The process by which adolescents go between commitment and in-depth exploration exemplifies what Erikson meant by a "crisis", because, while, this process is important for identity formation, it represents a state of confusion or "uncertainty" for the adolescent.

Some research has suggested developmental pathways towards identity formation. For example, younger adolescents are more likely to be in the diffusion or foreclosure status, while older adolescents are more likely to be in the moratorium or achievement status (Meeus, 2011). However, there is also evidence that identity formation may be more nuanced in that adolescents move in and out of identity statuses in a dynamic fashion (Schwartz et al., 2011). Furthermore, the movement between statuses may also vary depending on the domain of interest (e.g., occupation, peer relationships, gender, politics) (Goossens, 2001; van Doeselaar et al. 2018).

Importantly, the advent of the internet and online communication has provided adolescents with a unique avenue for identity exploration and expression. A Pew Research poll from 2015 reported that 95% of adolescents have access to a smart phone, 76% use social networking sites (e.g., Facebook, Snapchat, Instagram) and 45% reported

that they are online for most of the day (Lenhart, Duggan, Perrin, Stepler, Rainie & Parker, 2015). Adolescents are using this relatively novel way of communicating and sharing with others. Given that interaction with others is a primary means in which adolescents explore and confirm identity (Albarelo, Crocetti, & Rubini, 2018), adolescent activity in the digital space warrants investigation.

Indeed, there is increasing scientific investigation on the relationship between adolescent identity formation and the use of social media. Gray (2018) conducted an exploratory study investigating the reasons adolescents (grades 8-11) use social networking sites. Gray (2018) found that the most common reason was to chat with friends, followed by making plans with friends, sharing information about themselves, and viewing others' posts. This study provides evidence that the primary purpose of their activity is for interpersonal interaction. In addition, Cyr, Berman, and Smith (2015) investigated whether high school adolescents' frequency and preference for use of communication technology (e.g., texting, instant messaging, twitter and social networking) was related to identity statuses, identity distress, and existential anxiety. They found significant positive associations between usage, identity distress, and existential anxiety, but no significant relationship between usage and the identity statuses. They speculated that more communication technology usage may increase the distress associated with identity processes (e.g., exploration of commitments), but not affecting identity statuses. It could be that more social interaction increases the opportunity for identity exploration resulting in more uncertainty about oneself. Given that this study was correlational, it could also be that adolescents who already have a sense of uncertainty are more inclined towards online communication.

In another study, Kurek, Jose and Stuart (2017) conducted a latent profile analysis to determine whether adolescents' usage of information and communication technologies (ICT) (e.g., Facebook, surfing the internet, texting) was related to self-esteem, self-image, externalizing behaviors, and two aspects of identity presentation: false self and authentic self. The identity questions about false self and authenticity asked about adolescent's portrayal of themselves in general, not regarding the portrayal of themselves on social media platforms specifically. They found that adolescents in the elevated ICT use profile were more likely to report low self-esteem, low self-image, and were more likely to endorse portrayal of a false self and a less authentic self. Again, the directionality of these relationships cannot be determined based on these studies, however, they point to a relationship between online communication and identity factors.

The exploration of oneself in online contexts may also provide avenues through which adolescents can experiment with identity and connect with others. Youtube is a popular platform in which adolescents have been found to engage in identity exploration, both as content creators and viewers. For example, researchers conducted a qualitative analysis of the content of 22 Youtube channels that had at least 100,000 subscribers, 10,000 hits, and also contained some discussion of adolescent identity development (Pérez-Torres, Pastor-Ruiz & Abarrou-Ben-Boubaker, 2018). They found that the "Youtubers" spoke about many issues related to identity formation such as puberty, peer relationships, gender identity, and social relationships, with the goal of trying to relate to other adolescents' experiences. Viewers showed their support by commenting on the videos and expressing their own experiences with those aspects of identity formation.

This study presented evidence of a potentially positive way adolescents use digital platforms to communicate and connect with peers and explore their identities.

Digital technology has introduced a new medium for adolescents to explore their identities. This has opened up an important field of research, considering that adolescents use digital platforms for both communication with peers and individual expression, and the use of online platforms is related to multiple aspects of identity formation. Given that adolescents spend a considerable amount of their time on smart devices, the field would benefit greatly from studies that assess how adolescent identity development is reflected in the digital space. It has yet to be tested whether aspects of adolescent identity development have changed since such widespread digital technology use. There may be qualitative differences about how adolescents explore identity in the digital space versus how they explore identity in-person. For example, adolescents may move through the statuses of identity at a faster rate given that they have more opportunities for self-exploration and connection with others. Conversely, more opportunity for exploration might create more choices and could potentially slow down the processes that lead to identity achievement. The digital space may offer more opportunities for identity exploration, however, there may be consequences associated with the permanency of a digital footprint. The digital footprint creates a record of the “testing” of different identities that may lend itself to scrutiny by others. This might result in more cautious identity exploration related to potential online peer rejection or online bullying. However, the potential for anonymity in identity exploration online may offer more freedom and less social risk. In sum, these questions about adolescent identity formation in the digital space warrant examination.

Identity Formation and Mental Illness

Even though identity formation is associated with some experiences of distress, this is considered a normative aspect of identity development, and most adolescents make it through this stage of life without developing a mental disorder (Crocetti, et al., 2009). However, research has shown associations between adolescents who endorsed problems with identity formation and measures of adjustment problems (e.g., antisocial behaviors, depressed/anxious behaviors, hyperactive behaviors, peer problems/social withdrawal, and headstrong behaviors) related to both internalizing and externalizing disorders (Hernandez, Montgomery & Kurtines, 2006).

Identity Formation and Anxiety Disorders. Anxiety disorders are one of the most common mental disorders during adolescence (Ollendick, King & Muris, 2002). Research has shown anxiety disorders also tend to decrease over as the adolescent period ends (Hale III, Raaijmakers, Muris & Meeus, 2008). This falls in line with research on identity development, in that identity processes that are associated with increased levels of anxiety (e.g., reconsideration, moratorium) also tend to decrease over this period as well (Luyckx, Klimstra, Duriez, Van Petegem & Beyers, 2013). Additionally, higher levels of anxiety have been linked to the identity process of reconsideration specifically (Crocetti, Klimstra, Keijsers, Hale & Meeus, 2009). While reconsideration is a normative aspect of identity development, especially in early adolescence, older adolescents who experience higher levels of anxiety also tend to show higher levels of reconsideration and lower levels of commitment (Crocetti, et al., 2009). It could be that older adolescents who are struggling with making commitments compared to their same aged peers, may be at particular risk for anxiety. In addition, researchers have described an aspect of identity

formation associated with increased anxiety, termed ruminative exploration (Beyers & Luyckx, 2016). Like reconsideration, ruminative exploration is the process in which a person thinks about potential aspects of identity. However, ruminative exploration involves persistence in thinking about identity without making commitments (Beyers & Luyckx, 2016). This suggests that, while considering alternative identities is an important component of identity formation, too much deliberation without making a decision may be maladaptive.

Identity Formation and Depressive Disorders. The link between identity formation and depressive symptoms has been shown in many studies. For example, commitment making has been negatively associated with depressive symptoms, while ruminative exploration has been found to be positively associated with depressive symptoms (Luyckx, Klimstra, Duriez, Van Petegem & Beyers, 2013). In a large cross-sectional study looking at age trends in identity formation in people aged 14-30 years old, researchers found that depressive symptoms had a weaker positive association with ruminative exploration in mid adolescence and a stronger positive association for those in their late 20s (Luyckx, Klimstra, Duriez, Van Petegem & Beyers, 2013). It could be that ruminative exploration is less distressing during a time (mid adolescence) that is associated with more identity exploration in general. Ruminative exploration is related to exploration in breadth and depth, as it involves contemplation of potential options in regards to identity, but different in that rumination implies being “stuck” in a pattern of thought without moving towards commitment-making. In addition, Becht and colleagues (2017) examined within person, longitudinal associations between identity processes and depressive symptoms in order to determine directionality of change. Becht and colleagues

(2017) measured commitment, in-depth exploration, reconsideration, ruminative exploration and depressive symptoms of adolescents (M age=14.03) once a year for five years. They found that increases in uncertainty in regards to identity (i.e., increased reconsideration and ruminative exploration) predicted later increases in depressive symptoms. However, increases in depressive symptoms were not predictive of later identity uncertainty. This provides some evidence that the difficulty regarding identity formation appears to precede the development of depressive symptoms.

In another longitudinal study, researchers investigated whether strong identity commitments in adolescence would be predictive of less depressive symptoms in the context of stressful life events, and therefore could be considered a protective factor (van Doeselaar, Klimstra, Denissen, Branje & Meeus, 2017). They focused their investigation on career commitment (i.e., commitment to work and education) and interpersonal commitment (i.e., commitment to a best friend or romantic partner). They found that adolescents with weakening commitments over time were more likely to develop depressive symptoms. They also found that stronger commitments in the interpersonal domain, but not career domain, predicted decreases in depressive symptoms. However, they found that stronger commitments, in general, did not appear to buffer against depressive symptoms in the presence of stressful life events. This study highlights the importance of considering varying domains when measuring commitments, rather than measuring commitment in general.

Identity Formation and Suicidality. Suicidality, which can occur in the presence of depressive disorders, is an important experience to discuss in addition to depression in general, as it is one of the leading causes of death in adolescents 10-19 years old (Ruch et

al., 2019). Researchers have suggested that difficulty with the identity formation process may be a significant risk factor for adolescent suicidality (e.g., Bar-Joseph & Tzuriel, 1990; Ramgoon, Bachoo, Patel, & Paruk, 2006). In their study, Bar-Jospeh and Tzuriel (1990) found that suicidal adolescents scored lower than adolescents without mental illness or suicidality on components of ego identity related to self-control, meaningfulness, social recognition, and genuineness. The commitment and purposefulness component of ego identity was the only scale not significantly different between the two groups. Bar-Jospeh and Tzuriel (1990) suggested that having a more “consolidated ego identity” may serve to protect against suicidal ideation in regard to having a sense of self-control, meaningfulness, genuineness, and social recognition amidst the distress associated with identity formation in adolescence. It is worth considering that the adolescents with suicidality were lower on the self-control component of ego identity, as impulsiveness as a personality trait (opposite of self-control), has been found to be related to suicide attempts and completions in adolescents (Brent et al., 1994).

Portes and colleagues (2002) suggested that distress associated with identity formation, combined with stressors within the environment (e.g., lack of parental support or support outside the home, academic pressure, negative peer pressure), may put an adolescent at higher risk for suicidality. They proposed that the lack of consolidated identity further distorts adolescents’ perception of themselves and their future, and therefore, their ability to cope. Furthermore, adolescents struggling in many domains (e.g., parent relationships, peer relationships, academic) may be at higher risk for suicidal ideation or behaviors. Other researchers have theorized that the inability to establish a

sense of “sameness” throughout the changes associated with adolescent identity, across various domains, is related to suicidal ideation because of the inability to imagine a past, present and future self, both individually and in relation to others (Chandler, Lalonde, Sokol, Hallett & Marcia, 2003).

Another study investigated whether there were gender differences in identity confusion, coping strategies, and defense mechanisms in adolescents (12-17 years-old) who were admitted to the hospital after a suicide attempt (Foto-Ozdemir, Akdemir, & Cuhadaroglu-Cetin, 2016). The results indicated no significant gender differences in identity confusion. However, the sample, as a whole, had mean score one point away from the clinical cutoff for identity confusion, indicating problems with identity formation for many of these adolescents. Additionally, they found that adolescents who reported engaging in prior non-suicidal self-injury (NSSI) and who previously sought professional help were more likely to have endorsed higher levels of identity confusion.

In addition, research has shown that adolescents who identify as non-heterosexual and non-cisgender are more likely to experience suicidality (Guz, Kattari, Atteberry-Ash, Klemmer, Call & Kattari, 2021). Furthermore, adolescents who are questioning their identity in regard to gender and sexuality are at the highest risk for experiencing suicidality. While identity formation processes are associated with some level of distress for most adolescents, it appears that it may be particularly difficult for adolescents who identify with aspects of gender or sexual identities that are not wholly accepted within society.

Some research has indicated that there may be a “just right” point for identity formation processes and that too much, or too little at different stages of adolescence can

lead to maladaptive outcomes (i.e., psychopathology) (Beyers & Luyckx, 2016). Other researchers have suggested that it could be the experience of psychopathology that is then associated with problems in identity development (Klimstra & Denissen, 2017). In addition, Klimstra and Denissen (2017) argued that psychopathology may affect people's identity differentially. For example, consider Klimstra's and Denissen's example of two depressed youth; one who views their depression as central, or important, to their identity, while the other sees other aspects of their identity (e.g., academics, sports) as more central to their identity. The youth who has a strong sense of identity commitment in regard to their depression may not experience the positive outcomes that previous research has associated with high levels of commitment in general. Viewing aspects of their mental illness as central to their identity may be particularly problematic and dangerous for adolescents experiencing suicidal ideation and NSSI.

Klimstra and Denissen (2017) have discussed the importance of context when considering positive or negative identity development. In both cases, high levels of commitment are not, in and of themselves, a protective factor, as is often suggested. Marcia (2006) also warned that we need to be careful not to label, and potentially steer, the adolescent who appears to be on the track towards aligning with a negative identity (e.g., "self-injurer"). Social context and social influences are important factors to consider when investigating the relationship between psychopathology and identity (Klimstra & Denissen, 2017). Therefore, it is incomplete to make general statements about whether an identity status, in and of itself, is indicative of positive or negative risk/outcomes because the status is relative to the individual's context.

Natural Language Use, Psychopathology, and Identity Formation

The study of how humans use language offers insights into psychological and social phenomena. How we use words can convey information about psychological states. There are different approaches for analyzing written or spoke language, with some approaches highlighting the importance of considering the context of language use, while other approaches suggest that smaller units of language (e.g., word count, pronoun use, emotion words) convey rich information even when context is not taken into account (Pennebaker, Mehl & Niederhoffer, 2003). In fact, word count methods are one of the most widely used in research on natural language use. Pennebaker and colleagues (2003) discussed three psychological word count methods: (1) thematic content analysis which uses human judges to assess themes present in text, (2) pattern word analysis which he states is a "bottom-up" approach that is used to determine how similar bodies of text are to one another and (3) word count analysis which counts different types of words with goal of inferring psychological content. The development of computerized software to analyze language has allowed researchers to analyze large bodies of written or spoken word that have been collected in both laboratory (e.g., Hancock & Dunham, 2001) and naturalistic settings (e.g., Mehl & Pennebaker, 2003).

Language use is a rich source of information, not only because language is our primary means of communication and connection with the world around us, but also because research has shown that the way people use language is quite consistent, even when the topics vary. For example, Pennebaker and King (1999) investigated whether there were consistent individual differences in how people used language in their writing. Their study included writing samples from three groups of people and three different samples of writing. They collected writing from inpatients in a substance use program

who were instructed to write about a significant event of the day for 18 days, adult summer school students in a two-week health psychology course who were instructed to write about different topics for 20 minutes each day for 10 days (e.g., stream of consciousness, significant positive/negative childhood experience, reaction to relaxation exercise), and the most recent first author abstracts written by a number of prominent social psychologists. They used the Linguistic Inquiry and Word Count (LIWC) computer program to analyze language composition (e.g., word count, pronoun use), emotional, social and cognitive processes, and current concerns in each essay. They found that within person language components were largely consistent across over time, even when the topics varied, providing evidence that people have consistent patterns of language use and that language use analysis is a meaningful tool in understanding human behavior.

Natural Language Use and Psychopathology. Researchers have investigated the association between how people use words and mental illness, with particular attention paid to the relationship between language use and depression. There is a rich body of scientific literature that has examined the associations with psychopathology and personal pronoun usage (e.g., singular first-person [I, me, myself, mine], second person [you, your], third person [they, them, hers, his]). For example, one study analyzed word use in personal narrative essays written by college students who were depressed, formerly depressed, and never depressed in a laboratory setting (Rude, Gortner & Pennebaker, 2004). They found that college students who were currently depressed used more first-person-singular pronouns (specifically "I") in their essays compared to formerly and never depressed participants. They also found that currently depressed participants used

more negatively valenced words and less positively valenced words. This finding supports the theory that people who are depressed are more self-focused and that this self-focus is exemplified in their use of language (Watkins & Teasdale, 2004). The use of more negative and less positively valenced words also makes conceptual sense when considering depression.

The prolific use of digital technology to communicate with others has provided researchers with the opportunity to analyze language use “in the wild” and has led to a dramatic increase in the use of natural language processing (NLP) techniques to investigate potential links with mental illness. For example, a study analyzed the content from online mental health discussion forums that pertained to the following disorders: generalized anxiety disorder (GAD), borderline personality disorder (BPD), major depressive disorder, obsessive-compulsive disorder, and schizophrenia, and compared the content to an unrelated discussion forum (online personal finance forum) as a control (Lyons, Aksayli & Brewer, 2018). They used LIWC software to analyze the use of personal pronouns and emotion words on the discussion forums. Overall, they found that there was more use of first, second, and third person pronouns and negative emotion words on the mental health discussion forums compared to the control forum. In addition, they found that the people in the BPD forum used more third person-singular pronouns (e.g., she, him) than the control forum. People in the schizophrenia forum used the most third person-plural pronouns (e.g., they, them), although this association was only statistically significant when comparing word use in the schizophrenia forum to the word use in the GAD forum. The authors note that both of these findings are consistent with the conceptualization of some of the difficulties/symptoms of these disorders. One of the

common characteristics of BPD is difficulty with relationships and one of the common manifestations of schizophrenia is a focus on persecutory delusions perpetrated by others and feeling of being outside a social group. According to these findings, the focus on others is manifested in their use of third person pronouns.

In a similar study, researchers analyzed posts on anxiety related and non-anxiety related Reddit forums and were able to correctly differentiate between the two with 98% accuracy (Shen & Rudzicz, 2017). In an analysis of Twitter posts of those who self-disclosed anxiety disorders, a trained classifier was able to differentiate people with self-reported anxiety from a demographically controlled group with 79% accuracy (Dutta & De Choudhury, 2020). While these studies did not identify specific language use categories, they indicated that there were more general detectable differences in how people with anxiety use language.

Recently, researchers investigated whether there is a predictive relationship between online language use and symptoms of depression, generalized anxiety, and social anxiety (Stamatis et al. 2022). They analyzed the content of out-going text messages of adults and subsequent depressive, generalized anxiety, and social anxiety symptoms. They found that greater use of pronouns and words related to anger, sadness, and anxiety were positively associated with symptoms of depression, generalized anxiety, and social anxiety, overall. When controlling for the generalized and social anxiety symptoms, depressive symptoms were positively associated with sadness words and negatively associated with social process and affiliation words. The authors suggested that less use of socially affiliative words is consistent with the decrease in social interaction often observed amongst people with depression. Controlling for depressive

and social anxiety symptoms, generalized anxiety symptoms were positively associated with social process words. They speculated that people with generalized anxiety may show an increase in social engagement, but experience anxiety or fear in those situations, rather than experiencing the social interaction as unpleasant. Controlling for depressive and generalized anxiety symptoms, social anxiety symptoms were positively related to anger words, sexual words, swear words, and third-person pronoun use. Here the authors reasoned that feelings of anger may be protective against rejection in people with social anxiety and that some sexual and swear words overlap within the category of profanity and therefore are a manifestation of anger.

Another study analyzed Twitter posts of people who stated that they had attempted suicide, and who also included a date of when they attempted suicide (Coppersmith, Leary, Whyne & Wood, 2015). They analyzed the tweets they posted prior to the suicide attempt using LIWC categories and found that people who attempted suicide used more words, compared to a control group, that fell into the following LIWC categories: death, health, sad, I, they, sexual, filler, anger, and negative emotions. They also were able to train a classifier to detect people who reported attempting suicide from people who reported depression.

Using similar techniques, researchers analyzed outgoing texts messages during periods of suicidality and depression in a sample of older adolescents (college undergraduates, mean age=20.42 years) who endorsed a past suicide attempt (Nobles, Glenn, Kowsari, Teachman & Barnes, 2018). The researchers asked the participants to identify the date of the suicide attempt, periods of suicidal ideation, and periods of depression without suicidal ideation. They analyzed word count, emotional tone, and

linguistic features (e.g., first-person pronoun) to train a classifier to differentiate between periods of increased suicidal ideation and periods of depression with no reported suicidal ideation and were able to do so with 80% accuracy.

Additionally, there is evidence that words associated with absolutism (e.g., always, completely, never) are associated with internalizing disorders and suicidality (Adam-Troian & Arciszewski, 2020; Al-Mosaiwi & Johnstone, 2018). In an analysis of search trends on Google, researchers found a significant association between absolutist words and suicide rates in the United States (Adam-Troian & Arciszewski, 2020). Natural language use analysis of discussion forums pertaining to specific mental health problems (anxiety, depression, and suicidality) found that absolutist words were associated with mental illness discussion forums, compared to control (Al-Mosaiwi & Johnstone, 2018). Furthermore, the suicide discussion forums were associated with significantly more absolutist words compared to the anxiety and discussion forums. These findings provide evidence that absolutist words may be another important linguistic marker of internalizing problems and potentially indicative of the severity of mental illness (Al-Mosaiwi & Johnstone, 2018).

A recent study analyzed whether there were linguistic differences in spoken language amongst adolescents at high risk for psychosis, with and without suicidal ideation (Dobbs et al., 2023). They predicted there would be differences in use of words related to sadness, anger, stress, loneliness, and use of the word “I.” They found that adolescents considered high risk for psychosis who endorsed suicidal ideation used more words related to anger, compared to adolescents at high risk for psychosis with no

suicidal ideation and a control group. There were no differences detected in use of words related to sadness, stress, loneliness, or use of the word “I.”

A recent systematic review examined 31 studies that used NLP algorithmic methods in attempts to detect suicidality (Young, Bishop, Humphrey & Pavlacic, 2023). The studies reviewed used several different types of language data collected from a variety of sources including patient medical records, social media (e.g., Twitter, Reddit), text conversations between online therapist and client, and suicide notes. Measures of suicidality varied across studies as well and included samples of people who previously endorsed suicidality, as indicated in medical records, and people who later died by suicide. They found that regardless of specific algorithmic method, NLP methods, in general, were able to detect suicidality across varying types of language use data with a high level of precision.

Another group of researchers conducted a systematic review of studies that investigated the relationship between suicidal thoughts, suicidal behaviors, and specific linguistic features (Homan et al., 2022). They indicated there is some evidence that there may be linguistic differences between suicidal thoughts and suicidal behaviors. In a few studies they reviewed, researchers found that people who endorse suicidal thoughts were more likely to use language intensifiers (e.g., completely, really) and superlatives (e.g., biggest, fastest), while people who endorsed suicidal behaviors had greater overall word usage, greater use of first and second person pronouns, and greater use of nouns. Overall, however, Homan and colleagues found that the results were largely mixed in terms of making this distinction. Linguistic features such as death word use, first-person pronoun use, and negative emotionality were positively associated with both suicidal thoughts and

suicidal behaviors in many studies. They indicated one of the issues literature might be around distinguishing people who only have suicidal thoughts from people who have suicidal behaviors as well.

These studies provide evidence that there are detectable differences in how people use words and that they may vary as a function of specific disorder and illness symptomology. Not only have researchers capitalized on the huge corpuses of data available through medical records and social media platforms, but recent more work has utilized data collected directly from digital devices (i.e., smartphones). This is particularly relevant to adolescent populations, as digital devices are one of their primary means of communication.

In addition, it is important to understand how adolescents engage in the digital space. Researchers investigated adolescents' motivations, and potential detrimental effects, of social media use in a qualitative study using adolescent focus groups (Throuvala, Griffiths, Rennoldson & Kuss, 2019). Throuvala and colleagues (2019) found that adolescents are motivated to engage with social networking platforms in order to fulfill a need for controlling perceptions of themselves and maintaining relationships. They suggest that this could explain the potentially negative effects of social media use, described as fear-of-missing-out (fomo) and nomophobia (fear of being without smartphone). Another focus group study investigated processes related to how adolescents share images online (Bell, 2019). Bell (2019) found that adolescents spend a considerable amount of time curating their images before posting. Bell (2019) also found that one of the primary reasons for posting images was to maintain their relationships with people offline as well. This study is interesting in the context of studies of natural

language use online because it poses the question of whether adolescents may be curating their text responses in such as fashion as well. These studies also suggest that adolescents may use their devices differently than adults. This highlights the need for caution in assuming findings from studies with adults samples will be replicated in adolescent samples as well.

This area of research has utility beyond pure scientific exploration of language use in that, detecting language use patterns in people who are distressed in a passive, naturalistic environment, may aid the development of just in time adaptive interventions (JITAI) (Nahum-Shani et al., 2018). If language use patterns are predictive of periods of increase in mental illness symptoms severity and heightened risk for suicide, this creates the opportunity to intervene in the moment rather than relying on the person to initiate access to support from health professionals, crisis resources, social support systems, or wait until their next therapy appointment.

Natural Language Use and Identity Formation

The study of natural language use in relation to adolescent identity formation appears to be under-investigated. However, in the field of linguistic anthropology, researchers have suggested that identity is inextricably linked to the study of language and culture, because identity and language are fundamentally a cultural and relational phenomenon (Bucholtz & Hall, 2004). In addition, language use can be viewed as the outward expression of “internal mental states” (Bucholtz & Hall, 2010, p. 18), including how a person thinks of oneself. While linguistic anthropological study tends to focus on the role of language use on a societal or cultural level, it stands upon the foundation that individual minds collectively comprise society. Additionally, this highlights the

importance of language use in reference to the individual's experience, which is linked to how individuals think about themselves/interact with others.

Previous work on the study of natural language use in relation to mental disorders provides a good starting point into the investigation of natural language use and identity formation. For example, there is a rich literature on the positive association between first-person pronoun use and internalizing disorders hypothesized to be related to the self-focused nature of these disorders (e.g., Luyckx et al., 2013). It could be that adolescents who are in the moratorium (i.e., searching) status may use more first-person pronouns, reflecting a more intense focus on the self, while adolescents who are in the commitment status may use less first-person pronouns, representing a more solidified sense of self, and therefore less self-focus.

In addition, research on the associations between natural language use and psychological states offers clues as to what sort of language use may be associated with these identity formation processes. By utilizing existing research on natural language use and indices of well-being and psychopathology, we can begin to test hypotheses about potential relationships with identity formation. Research has indicated that identity consolidation may vary depending on the domain. For example, an adolescent who is academically driven and passionate about their career trajectory might fall in the commitment stage of identity formation in regard to career. However, this same adolescent may be unsure about how they want to express their gender identity and fall in the moratorium or diffusion stage of identity formation for gender. They may also be questioning the religious beliefs taught to them by their parents and be exploring other religious ideologies and engaged in in-depth exploration. Domain specific language use

could provide information about what areas of the adolescents' experience are the most salient in the context of mental illness and suicidality.

The Current Study

This study explored the relationship between psychopathology, suicidal thoughts and behaviors (STBs), adolescent identity formation, and natural language use utilizing text data collected from digital devices with the aim of identifying potential markers for psychopathological risk states.

Research has shown that there are differences in the ways people with psychopathology use language. NLP techniques have identified differences in language use between control group samples and people with anxiety disorders (Dutta & De Choudhury, 2020; Shen & Rudzicz, 2017) and depressive disorders (Nobles et. al., 2018; Rude, Gortner & Pennebaker, 2004; Watkins & Teasdale, 2004). In addition, some research suggests that there is variation in psycholinguistic features *between* mental illness categories as well. This suggests that differences in psychopathology may extend beyond the manifestation of a mental illness in general. In other words, there may be key linguistic features of language use that coincide with specific mental illnesses (e.g., anxiety disorder versus depressive disorder; depressive episode with suicidal ideation/behavior versus depressive episode with suicidal ideation/behavior). To my knowledge, no study has assessed whether differences in language use is detectable *between* mental illness categories and within those with or without suicidal thoughts and behavior (STB).

In addition, this study attempts to fill in the gaps in the literature surrounding the relationship between natural language use, identity formation, and mental illness during

adolescence. Given that psychopathology is associated with linguistic features and identity distress, I propose that identity distress will be detectable within natural language use among adolescents with psychopathology.

Because identity distress has been theorized to be a central component to psychopathology for some adolescents, it may be difficult to separate distress in general from distress related to identity specifically. I proposed that utilization of domain specificity regarding defining distress pertaining to identity will provide insight into such differences. For example, words that may be associated with identity (singular personal pronoun use (e.g., I, me), negative affect (potential indicator of distress), and words related to specific domains that are central to the adolescent experience (e.g., academics, relationships, gender, sexuality) could be indicators of distress as it is related to identity formation.

Specific Aims

The overarching goal of this study is to utilize adolescents' natural digital social communication to identify linguistic markers of STBs, depression, anxiety, and identity distress in adolescents.

Aim 1. To examine whether there are group differences in language use amongst adolescents who have psychopathology without STBs, adolescents with STBs without prior suicide attempt(s), and adolescents with STBs and prior suicide attempts. Previous research has provided evidence that there might be differences in language use between people with mental illness, in general, and people who have experienced STBs (e.g., Al-Mosaiwi & Johnstone, 2018; Coppersmith et al., 2015; Nobles et al., 2018). Language use related to the use of first-person pronoun use (Coppersmith et al., 2015; Nobles et al.,

2018), absolutist words (e.g., always, never) (Adam-Troian & Arciszewski, 2020; Al-Mosaiwi & Johnstone, 2018), negative emotion (e.g., anger, sadness) (Coppersmith et al., 2015; Nobles et al., 2018), and death (Coppersmith et al., 2015) have been shown to have some specificity.

Hypothesis 1: I predict that there will be group differences in language use between adolescents with psychopathology, but without STBs (psychiatric control), adolescents with STBs and no suicide attempts (ideators), and adolescents with STBs and prior suicide attempt(s) (attempters) for the following language use categories: first-person pronoun use (e.g., I, me), sadness, anger, death, and absolutist words (e.g., never, always).

Testing Hypothesis 1: Multivariate analysis of variance (MANOVA) will be conducted to test for group differences (psychiatric control, ideators, attempters) in language use variable listed above.

Hypothesis 1a: These relationships will remain significant after controlling for depressive symptoms, therefore demonstrating some specificity of these associations to STBs.

Testing Hypothesis 1a: Multivariate analysis of covariance (MANCOVA) will be conducted to test for group differences (psychiatric control, ideators, attempters) in language use variable listed above, controlling for depressive symptoms, in order to isolate the effect of STBs from depressive symptoms on language use.

Aim 2. To investigate the association between the quantitative dimensions of anxiety, depression, and STBs and language use. This aim will include words previously associated with STBs and depression (see Aim 1), and also include investigation of words that may be related to anxiety. Some research has indicated that anxiety symptomology is

associated with higher use of second/third person pronouns and words related to anxiety (e.g., Sonnenschein, Hofmann, Ziegelmayer, & Lutz, 2018).

Hypothesis 2a: I predict that the severity of self-reported suicidal thoughts and behaviors will be associated with greater use of words from the following language use categories first-person pronoun use, sadness, anger, death, and absolutist words.

Testing for Hypotheses 2a. I will conduct multiple regression analysis to test the relationship between self-reported STB severity and first-person pronoun use, sadness, anger, anxiety, death, and absolutist words.

Hypothesis 2b: I predict that the severity of self-reported depressive symptoms will be associated with greater use of words from the following language use categories: first-person pronoun use, sadness, anger, death, and absolutist words.

Testing for Hypotheses 2b. I will conduct multiple regression analysis to test the relationship between self-reported depression severity and first-person pronoun use, sadness, anger, death, and absolutist words.

Hypothesis 2c: I predict that the severity of self-reported suicidal thoughts and behaviors will be associated with greater use of words from the following language use categories first-person pronoun use, sadness, anger, death, and absolutist words, controlling for self-reported depressive symptoms.

Testing for Hypothesis 2c: I will conduct multiple regression analysis to test the relationship between self-reported STB severity and first-person pronoun use, sadness, anger, death, and absolutist words, controlling for self-reported depressive symptoms.

Hypothesis 2d: I predict that the severity of self-reported anxiety symptoms will be associated with words associated with the following language use categories: second- and third-person pronouns (e.g., you, she, he, they), anxiety, and health.

Testing for Hypotheses 2d. I will conduct multiple regression analysis to test the relationship between self-reported anxiety severity and second/third-person pronouns, anxiety, and health word usage.

Aim 3 (Exploratory): To investigate whether patterns of online natural language use can detect identity distress among adolescents experiencing psychopathology utilizing concepts of domain specificity and identity formation. Although I did not have a direct psychometric measure of identity stress in the study, I utilized theory from identity development to include words related to domains that have been shown to be important in adolescent identity development (e.g., friend, family, work, religion, health, and sex). These identity domains, in conjunction with words related to self-focus (first-person pronouns) or other-focused (e.g., second/third person pronouns) and distress (e.g., negative emotions) may provide insight into areas of difficulty for adolescents who experience higher amounts STBs, depression, or anxiety. Given the exploratory nature and novelty of the following research questions, I will conduct stepwise regressions to examine these relationships.

Research Question 1: Is the severity of STB symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

Testing for Question 1. I will conduct a stepwise regression analysis to test the relationship between STB severity and first, second/third person pronouns, sadness, anger, anxiety, friend, family, work, religion, health, and sex.

Research Question 2: Is the severity of depressive symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

Testing Question 2. I will conduct a stepwise regression analysis to test the relationship between depression severity and first, second/third person pronouns, sadness, anger, anxiety, friend, family, work, religion, health, and sex.

Research Question 3: Is the severity of anxiety symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

Testing Question 3. I will conduct a stepwise regression analysis to test the relationship between anxiety severity and first, second/third person pronouns, sadness, anger, anxiety, friend, family, work, religion, health, and sex.

CHAPTER II

METHOD

Participants

Participants were high risk adolescents between 13-19 years old (M age=16.94) who were either (a) recent suicide attempters with current ideation (n=30), (b) current suicide ideators with no attempt history (n=45), and (c) a psychiatric control group with no STB history (n=37) (total n=112).

Eighty-six (76.8%) participants identified as female, 20 (17.9%) identified as male, four (3.6%) identified as transgender male, and two (1.8%) identified as transgender female. Eighty-eight participants (78.6%) identified as having non-Hispanic ethnicity and 24 (21.4%) identified as having Hispanic ethnicity. Fifty-five (49.1%) participants identified as White, 16 (14.3%) identified as Black or African American, 14 (12.5%) identified as more than one race, 11 (9.8%) identified as unknown or not reported, 15 (13.4%) identified as Asian, and one (.9%) identified as Native American or Alaskan Indian.

Procedure

Participant Recruitment. Participants for this study were recruited from inpatient and outpatient clinics by staff from New York State Psychiatric Institute/Columbia University Irving Medical Center and the University of Pittsburgh Medical Center.

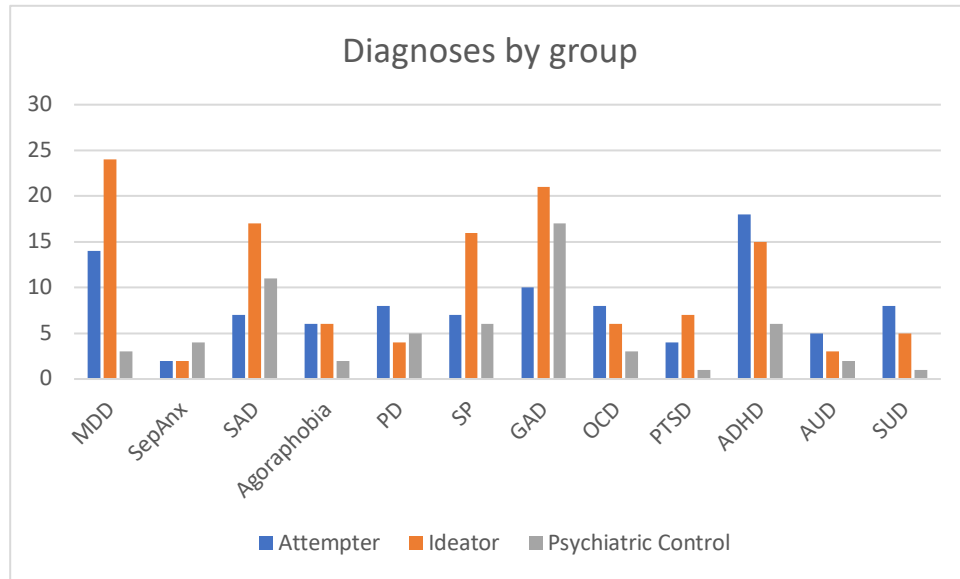
Consent. All participants provided written informed assent. Written informed consent was provided by the participants' parent/legal guardian.

Inclusion/Exclusion Criteria. Participants were included in the *suicide attempt group* if they were 13-18 years-old, speak English, had a parent who speaks English or Spanish, and had an Android or iPhone7+ device, met criteria for a current mood/anxiety or substance use disorder, endorsed current STB, and endorsed suicide attempt in the past year. Participants were excluded from the *suicide attempt group* if they endorsed imminent intent to kill oneself, were actively manic or psychotic, or met criteria for autism spectrum disorder. Participants were included in the *suicide ideation group* if they were 13-18 years-old, spoke English, had a parent who spoke English or Spanish, and had an Android or iPhone7+ device, met criteria for a current mood/anxiety or substance use disorder, and endorsed current STB. Participants were excluded from the *suicide ideation group* if they endorsed prior suicide attempt, endorsed imminent intent to kill oneself, were actively manic or psychotic, or met criteria for autism spectrum disorder. Participants were included in the *psychiatric control group* if they were 13-18 years-old, spoke English, had a parent who spoke English or Spanish, and had an Android or iPhone7+ device, met criteria for a current mood/anxiety or substance use disorder, denied current STB. Participants were excluded from the *psychiatric control group* if they endorsed prior STB, endorsed prior suicide attempt, endorsed imminent intent to kill oneself, were actively manic or psychotic, or met criteria for autism spectrum disorder.

Materials and Measures

Baseline measures. MINI-KID. Assessment of mental illness with the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID; Sheehan et al., 2010). (Figure 1.).

Figure 1. Diagnoses by Group



Abbreviations: Generalized anxiety disorder (GAD), major depressive disorder (MDD), attention deficit hyperactivity disorder (ADHD), social anxiety disorder (SAD), specific phobia (SP), panic disorder (PD), obsessive compulsive disorder (OCD), substance use disorder (SUD), posttraumatic stress disorder (PTSD), alcohol use disorder (AUD), separation anxiety disorder (SepAnx), oppositional defiant disorder (ODD).

SITBI. The Self-Injurious Thoughts and Behaviors Instrument (SITBI; Nock et al., 2007) structured clinical interview was used to assess suicide attempts, suicide gestures, and non-suicidal self injury.

SSI. Suicidal thoughts and behaviors were assessed with Beck’s Scale for Suicide Ideation self-report questionnaire (SSI, Beck, Kovacs, Wiessman, 1979). The SSI is a 19-item self-report questionnaire that assesses whether suicidal thoughts have been present over the previous week and how intense they were. Scores range from 0-38, with higher scores indicating greater STBs. (See Appendix C.)

MFQ. Depressive symptoms were assessed with the Mood and Feelings Questionnaire (MFQ; depressive symptoms; Angold et al., 1995). The MFQ is a 13-item self-report questionnaire that assesses depressive symptoms experienced over the past two weeks. Scores range from 0-26, with scores over 12 indicating greater severity in depressive symptoms. (See Appendix D.)

SCARED. Anxiety symptoms were assessed with the Screen for Anxiety Related Emotional Disorders short form (SCARED; anxiety symptoms; Birmaher et al., 1999). The SCARED short form includes five questions that assess anxiety symptoms severity. Scores can range from 0-10, with higher scores indicating greater anxiety symptoms. (See Appendix E.)

Intensive longitudinal monitoring language measure. EARS. The Effortless Assessment of Risk States (EARS) is a HIPAA-compliant software tool developed to collect data via an app downloaded to participants' personal smartphones. The EARS tool collects active response data (i.e., custom questionnaires pushed to the phone) as well as passive data on language through normal phone usage. (The EARS tool passively collects other data (e.g., facial expression, geographic location) that were not used in the current study.) Baseline data were collected during the initial 2 weeks of use and continuous follow-up data were collected during a 6 month follow-up period. The EARS tool collects all language typed into the phone via a custom-built keyboard. EARS passive data do not require any user input, rather it runs in the background of a participant's smartphone. Participants were compensated \$300 for providing 6-months passive sensing data (payments issued in \$100 increments and compliance is determined at the 1-, 3-, and 6-

month follow-up assessment; payments issued electronically via an RFMH ClinCard or via an UPMC Vincent Card).

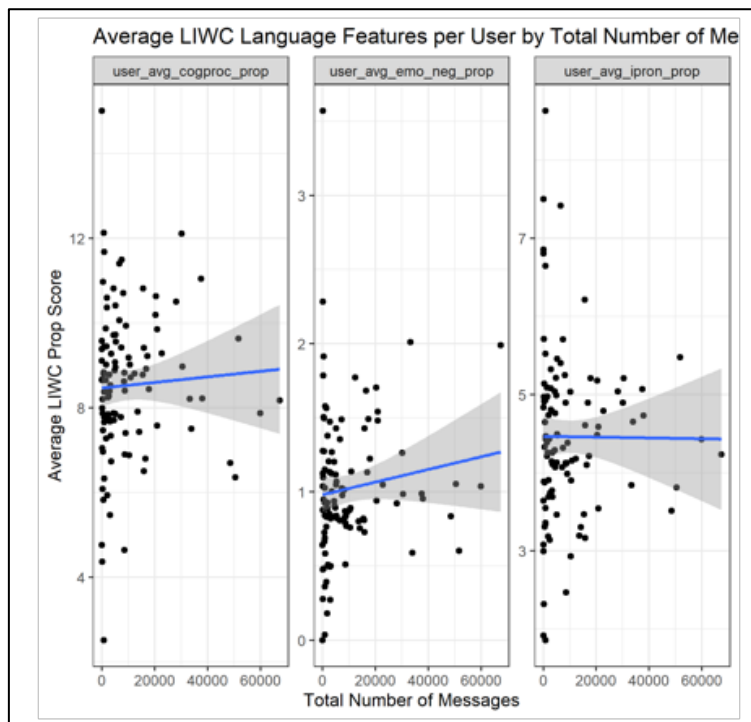
Language variable creation. Language data were processed using the Linguistic Inquiry and Word Count (LIWC) 2022 software (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022). The LIWC software calculated the proportion of words per message for the following language categories: pronouns, cognitive processes, affect, social relationships, physical, and lifestyle. Next, average proportions were calculated to produce a single estimate per linguistic category, per participant.

Additionally, the initial sample included 140 adolescents who had any amount of text data. The range in number of texts from each participant was between 13 and 67,295 text messages. The LIWC software used to analyze text computed mean proportion values for word categories. In effort to avoid over- or under-inflating the value of proportion of word use, and more importantly, the interpretation and inferences made of a given participant's language use, a cut-off for minimum number of text messages was created, prior to examination of assumptions. Visual inspection of scatterplots was conducted with the average proportion for a sample of LIWC categories (cognitive processes, negative emotion, and first-person pronoun use). Cognitive processes and negative emotion categories were selected because they subsume the language variables of interest in this study. First-person pronoun use category was selected because it is one of the most robust language variables used to investigate relationships with mental illness. Visual inspection of the scatterplots shows a large spread in proportion value (y-axis) for participants with very few messages (x-axis) (Figure 2). The proportion values

appear to centralize around the 500 messages mark, therefore participants with at least 500 messages were included in the study (n=112 included, n=38 excluded).

Figure 2

Scatterplot Total Number of Messages



Note. Cognitive processing language category, negative emotion language category, and first-person pronoun language category.

Data Analysis

Analyses Addressing Aims. The data were analyzed using IBM SPSS Statistics (Version 23).

This study utilized linear regression and linear mixed-effects models to test the hypotheses. These statistical tests rely on the assumptions that data are independent and normally distributed. Therefore, data visualization and statistical tests of normality were conducted prior to analyses. Research has indicated that checking violations of

assumptions is, unfortunately, an uncommon practice amongst researchers, and may result in inappropriate test selection or increase in the likelihood of Type I or Type II errors (Hoekstra, Kiers & Johnson, 2012). Additionally, given the novelty of the text data used in this study, the determination to check these assumptions were particularly warranted. Investigation of independence, outliers, distribution and normality, and multicollinearity were reviewed prior to completing statistical analyses. Variance and homoscedasticity assumptions were tested and interpreted in conjunction with the statistical analyses testing the hypotheses.

Independence. The participants in this study were recruited via advertisements placed in outpatient and hospital settings. Targeted recruitment was utilized because the population of interest was adolescents with psychiatric disorders and suicidal ideation. While not a true random sample from the population of interest, the unique vulnerability of this population justifies the sampling method. The observations of data are unique to each individual participant. Therefore, the independence of data is assumed.

Normality. The assumption of normality of data is necessary to utilize the statistical analyses required to test this study's hypotheses. Normality of data is necessary when analyses utilize the theory of regression to the mean, which assumes a normal distribution of data. There are myriad ways to assess normality in a given data set, including both visual representation of data, as well as statistical tests. However, given the novelty, particularly of the language data, I felt it was justified to utilize multiple means of assessment. Therefore, I used histograms, box and whisker plots, and quantile quantile (QQ) plots for visualization and inspection of qualitative nature of data.

Additionally, I conducted Lilliefors corrected Kolmogorov-Smirnov and Shapiro-Wilk tests to provide quantitative assessments of normality.

Histogram. In order to view whether the assumption of normal distribution is met, I created histograms of dependent variables (depression symptoms (MFQ), anxiety symptoms (SCARED), and suicidal ideation symptoms (SSI) and the distribution of average proportion of messages by category created by LIWC for the independent variables (first-person pronouns, second-person pronouns, anxiety, anger, sadness, death, absolutistic, family, friend, work, religion, health, sexual) (see appendix A, Figure 1). Visual inspection of first-person pronoun, second person pronoun, anxiety, anger, absolutistic, and work words are generally normally distributed. Visual inspection of the MFQ and SCARED variables indicated the data are normally distributed with a somewhat uniform shape. Visual inspection of suicidal ideation symptom variable is positively skewed due to the large number of “zero” (i.e., no suicidal ideation) responses by participants. In addition, the sadness, death, family, friend, religion, and sexual words variables generally had a positive skew as well.

Box and Whisker Plot. Next, I created box and whisker plots for further investigation of symmetry of data and potential outliers (see Appendix A, Figure 2). When the whiskers of a box and whisker plot are relatively equal on each side of the box, it suggests the data are distributed symmetrically. In addition, the box and whisker plots identified potential outliers for each variable. Visual inspection of box and whisker plots indicated symmetry for the MFQ and SCARED variables. The SSI variable and language use variables indicated positive skew likely driven by potential outliers. In addition, there

were extreme outliers detected for SSI and first-person pronouns, family, work, and religion language use variables.

QQ Plot. Next, I created normal QQ plots to check the assumption data are normally distributed (see appendix A, Figure 3). Normal QQ plots plot theoretical normal distribution quantiles (y-axis) against the quantiles in a sample (x-axis) (Marden, 2004). Data in the sample are determined to be normally distributed when they are close to the reference line (i.e., expected distribution for normality). Additionally, potential outliers are suggested by deviance from the reference line. Review of the MFQ, SCARED, first-person pronoun, second-person pronoun, anxiety, anger, absolutist, and work variables indicated the data are generally normally distributed, with some deviations from the reference line. Review of the SSI QQ plot indicated the data are non-normal and heavily skewed. Review of the sadness, death, family, friend, religion, health, and sexual variables indicated skewness and potential non-normality as well. Every normal QQ plot had deviations from the reference line, indicating potential outliers as well.

Kolmogorov-Smirnov and Shapiro-Wilk Tests. Next, I conducted Lilliefors corrected Kolmogorov-Smirnov and Shapiro-Wilk tests for quantitative assessment of normality. Results from the Kolmogorov-Smirnov test indicate normality ($p < .05$) for MFQ, second person pronoun, and anger and non-normality ($p > .05$) for the remaining variables (Table 1). Results from the Shapiro-Wilk test indicated normality for only second person pronoun variable ($p < .05$).

Table 1*Lilliefors corrected Kolmogorov-Smirnov (K-S) and Shapiro-Wilk Tests*

<u>Self report variable</u>	<u>K-S (p)</u>	<u>Shapiro-Wilk (p)</u>
MFQ	.071	.971*
SSI	.195*	.828*
SCARED	.104*	.966*
<u>Language variable</u>		
FPP	.094*	.930*
SPP	.073	.982
Anxiety	.099*	.967*
Anger	.063	.971*
Sad	.096*	.947*
Death	.102*	.910*
Absolutistic	.109*	.958*
Family	.109*	.808*
Friend	.174*	.913*
Work	.087*	.958*
Religion	.116*	.862*
Health	.102*	.942*
Sex	.118*	.897*

Note. MFQ = Mood and Feelings Questionnaire, SSI = Beck's Scale for Suicide Ideation, SCARED = Screen for Anxiety Related Emotional Disorders, FPP = first-person pronouns, SPP = second-person pronouns. * $p < .05$.

Summary and Action Taken. Taken together, review of histograms, box-and-whisker plots, normal QQ-plots, and Lilliefors corrected Kolmogorov-Smirnov and Shapiro-Wilk tests indicated potential violations of the normality assumption. Therefore, steps were taken to address these issues including removal of outliers and transformation of variables.

First, I removed the extreme outliers detected in the box and whisker plots. No extreme outliers were detected for the dependent variables (SSI, MFQ, SCARED). There were extreme outliers detected for the following language use variables: first-person pronouns (n=1), family (n=3), work (n=1), religion (n=1), and health (n=1). Rather than remove extreme outliers from the sample completely, scientists suggest Winsorization of values to mitigate the effect of outliers (Dixon, 1980; Vijendra & Shivani, 2014). Therefore, I transformed each extreme value to the next highest value that was not an outlier.

Next, I conducted square root transformation on variables where potential non-normality was present. Previous research suggests computing square root transformation, versus logarithmic, for data with positive values and can also be used on data with zero values, such as the SSI variable used in this study (Osborne, 2002). The SSI variable was positively skewed, due to the large number of zero values (indicating no suicidal ideation). Previous research has come across the issue of positive skew with the Beck's Scale for Suicide Ideation as well and have consistently used data transformation methods to help with normality assumption (e.g., Holi et al., 2005). Additionally, the language use variables generally showed a positive skew. Therefore, square root

transformation was conducted for the SSI variable and the language use variables in order to stabilize the variance.

After first Winsorizing extreme outliers and then computing square root transformation on the aforementioned variables, I created histograms (Appendix A, Figure 4), QQ plots (Appendix A, Figure 5), and ran Kolmogorov-Smirnov and Shapiro-Wilk tests (Appendix B, Table 1) with the transformed variables in order to view the effects of data transformation. Qualitative visual inspection and quantitative assessment of transformation indicated that removal of outliers and square root transformation increased normality of data. In some instances, the Kolmogorov-Smirnov and Shapiro-Wilk tests remained significant, indicating potential for non-normality. Research has shown that normality tests are sensitive to small changes in large sample sizes, therefore, significant results do not necessarily suggest violation of this assumption (Ghasemi & Zahediasl, 2012). As stated previously, Shapiro-Wilk tests are sensitive to larger samples and significance does not assume violation of normality, nor does it indicate that analysis of variance to be inappropriate for the given data (Ghasemi & Zahediasl, 2012). Additionally, recent research indicates that linear mixed-effects models, such as those used in this study, are generally robust when data potentially violate assumptions of normality (Knief & Forstmeier, 2021). Taken together, Winsorization and square root transformation have improved the distributions towards normality, therefore, the following analyses were conducted utilizing the transformed values with the assumption of normality being met.

CHAPTER III

Results

Descriptive Statistics

Descriptive statistics for suicidal ideation (SSI), depressive symptoms (MFQ), anxiety symptoms (SCARED), and LIWC language variables are presented below in Table 3.

Table 3

Sample Means and Standard Deviations

<u>Self-report Variable</u>	<u>M</u>	<u>SD</u>
MFQ	12.29	6.53
SSI	2.03	1.78
SCARED	4.25	2.32
<u>Language Variable</u>		
FPP	2.33	.20
SPP	1.44	.15
Anxiety	.37	.09
Anger	.45	.11
Sad	.41	.13
Death	.37	.12
Absolutistic	1.28	.14
Family	.74	.19
Friend	.54	.15

Table 3, continued

Work	.94	.20
Religion	.47	.18
Health	.65	.13
Sex	.43	.19

Note. MFQ = Mood and Feelings Questionnaire, SSI = Beck's Scale for Suicide Ideation, SCARED = Screen for Anxiety Related Emotional Disorders, WC= word count, FPP = first-person pronouns, SPP = second-person pronouns.

Descriptive statistics by group (psychiatric control group, ideator group, attempter group) suicidal ideation (SSI), depressive symptoms (MFQ), anxiety symptoms (SCARED), and language variables are presented in Table 4.

Table 4*Sample Means and Standard Deviations by Group*

	Psychiatric Control	Ideators	Attempters
Self-report variable	<i>M (SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>
MFQ	6.38 (4.38)	15.13 (5.49)	15.41 (5.09)
SSI	.26 (.59)	2.59 (1.40)	3.41 (1.49)
SCARED	3.86 (2.34)	4.60 (2.33)	4.21 (2.27)
<u>Language variable</u>			
FPP	2.33 (.21)	2.36 (.19)	2.26 (.21)
SPP	1.43 (.16)	1.43 (.15)	1.47 (.14)
Anxiety	.36 (.10)	.38 (.10)	.39 (.08)

Table 4, continued

Anger	.45 (.11)	.48 (.12)	.43 (.10)
Sad	.42 (.14)	.41 (.13)	.41 (.12)
Death	.33 (.12)	.37 (.12)	.41 (.11)
Absolutistic	1.29 (.15)	1.29 (.16)	1.23 (.10)
Family	.72 (.16)	.71 (.20)	.81 (.20)
Friend	.52 (.16)	.53 (.15)	.57 (.15)
Work	1.01 (.17)	.91 (.20)	.88 (.22)
Religion	.44 (.11)	.47 (.21)	.49 (.18)
Health	.61 (.11)	.66 (.13)	.69 (.15)
Sex	.40 (.17)	.40 (.21)	.45 (.17)

Note. MFQ = Mood and Feelings Questionnaire, SSI = Beck's Scale for Suicide Ideation, SCARED = Screen for Anxiety Related Emotional Disorders, WC= word count, FPP = first-person pronouns, SPP = second-person pronouns.

Aim 1

Multivariate analysis of variance (MANOVA) was used to examine between group differences (psychiatric control, ideators, attempters) and the use of first-person pronoun use, sadness, anger, death, and absolutist words (*Hyp 1*). The *Box's M* (Table 5) was not statistically significant, indicating the covariance matrices between groups are equal. The Levene's test was not significant for any of the dependent variables indicating equality of error variances (Table 6). Therefore, the assumption of homogeneity of variances is met.

Table 5*Box's M Test for Homogeneity of Variance Matrices*

Box's M Statistic	F	df1	df2	p
31.98	.99	30	30,699.96	.48

Table 6*Levene's Tests of Equality of Error Variances*

Dependent Variables	F	df1	df2	p
First-Person Pronoun	.30	2	109	.74
Sadness	.87	2	109	.42
Anger	.59	2	109	.56
Death	.12	2	109	.89
Absolutistic	2.64	2	109	.08

Results indicated the main effect of group was not statistically significant ($F(10, 210)=1.69$, Wilk's $\Lambda=.86$, partial $\eta^2=.08$, $p>.05$). Given that this result is nonsignificant, further interpretation of between subjects effects and *post hoc* tests are typically not warranted. However, the p -value was close to significance (.08), and given the novelty of the data, I wanted to ensure a thorough examination of the statistical results, in order identify potentially viable hypotheses for further research. Between subjects analyses revealed a significant effect for the use of words in the death category (Table 7).

However, first-person singular pronouns, sad, anger, and absolutist thinking word use showed no between group differences (Table 7).

Table 7

Between Subjects Effects by Group

		Mean Square	F	df1	df2	p
Group	First-Person Pronoun	.09	2.31	2	109	.10
	Sadness	.02	.03	2	109	.97
	Anger	.00	1.48	2	109	.23
	Death	.06	4.20	2	109	.02*
	Absolutistic	.03	1.47	2	109	.24

Note. * $p < 0.05$, ** $p < .01$.

Scheffe *post hoc* analysis showed that the significant effect for words related to death was between the psychiatric control group ($M=.33$, $SD=.12$) and the attempters group ($M=.41$, $SD=.11$), such that attempters used significantly more proportion of death words than the psychiatric control group ($Mdiff=.09$, $SE=.03$, $p < .05$).

Next, MANCOVA analysis was used to examine the above relationship, controlling for depressive symptoms (*Hyp 1a*). There was a statistically significant main effect of depression symptoms ($F(5, 103)=2.60$, Wilk's $\Lambda=.89$, partial $\eta^2=.11$, $p < .05$). Between subjects effects for depression symptoms were statistically significant for anger and sadness word usage, such that participants who scored higher on self-reported depression symptoms used more words related to anger and sadness (Table 8).

Table 8*Between Subjects Effects for Depression Symptoms*

		Mean Square	F	df1	df2	p
MFQ	First-Person Pronoun	.03	.85	1	108	.36
	Sadness	.02	5.19	1	108	.03*
	Anger	.05	4.50	1	108	.04*
	Death	.04	3.01	1	108	.09
	Absolutistic	.04	2.15	1	108	.15

Note. * $p < 0.05$, ** $p < .01$.

The main effect of group was not statistically significant when including depression symptoms in the model ($F(10, 206) = 1.67$, Wilk's $\Lambda = .86$, partial $\eta^2 = .08$, $p > .05$). There were no significant between subjects effects, nor significant pairwise comparisons.

Aim 2

A multiple regression analysis tested the relationship between self-reported STB severity (SSI) and first-person pronoun, sadness, anger, death, and absolutist word usage (*Hyp 2a*). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.11 or below), indicating that multicollinearity is low (Lavery, Acharya, Sivo & Xu, 2019). Next, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 6). Therefore, it was concluded that the assumption of homoscedasticity was met.

Results from the multiple regression analysis indicated that the language use variables did not significantly predict STB severity ($F(5, 105)=1.37, p>.05, R^2=.06$). The individual predictors were examined and showed that death word usage was a significant predictor of suicidal ideation severity, however, first-person pronoun, sadness, anger, and absolutistic word usage were not (Table 9). Participants who scored higher on self-reported STB symptoms used more words related to death.

Table 9

Results of Multiple Regression Individual Predictors for Predicting STB Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	1.08	2.44		[-3.76, 5.92]	.44	.66
First-Person Pronoun	-.27	.86	-.03	[-1.98, 1.44]	-.31	.76
Sadness	-.32	1.35	-.02	[-3.00, 2.36]	-.24	.81
Anger	-.56	1.64	-.34	[-3.18, 2.69]	-.34	.74
Death	3.58	1.38	.25	[.85, 6.32]	2.60	.01*
Absolutistic	.51	1.20	.04	[-1.86, 2.88]	.43	.67

Note. * $p<0.05$, ** $p<.01$.

Next, multiple regression analysis tested the relationship between self-reported depression severity (MFQ) and first-person pronoun use, sadness, anger, anxiety, death, and absolutist word usage (*Hyp 2b*). First, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 7). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.18 or below), indicating that multicollinearity is low. Therefore, it was concluded that the assumption of

homoscedasticity is met. The results from the multiple regression analysis indicated that the language use variables significantly predicted depression symptom severity ($F(5, 105)=2.71, p<.05, R^2=.11$). The individual predictors were examined and showed that death word usage was a significant predictor of depression symptom severity, however, first-person pronoun, sadness, anger, and absolutistic word usage were not (Table 10). Participants who scored higher on self-report measure of depression symptoms used more words related to death.

Table 10

Results of Multiple Regression Individual Predictors for Depressive Symptom Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	7.89	8.71		[-9.39, 25.17]	.91	.37
First-Person Pronoun	-3.07	3.08	-.11	[-9.81, 2.43]	-1.20	.23
Sadness	4.56	4.83	.09	[-5.02, 14.14]	.95	.35
Anger	7.71	5.86	.13	[-3.91, 19.34]	1.31	.19
Death	13.09	4.93	.25	[3.32, 22.85]	2.66	.00**
Absolutistic	2.03	4.27	.05	[-6.26, 10.66]	.52	.61

Note. * $p<0.05$, ** $p<.01$.

Next, multiple regression analysis tested the relationship between self-reported STB severity (SSI) and first-person pronoun use, sadness, anger, anxiety, death, and absolutist words, controlling for self-reported depressive symptoms (MFQ) (*Hyp 2c*). First, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 8). The variance inflation factor statistic (VIF) indicated that all dependent variables had a

tolerable VIF (1.20 or below), indicating that multicollinearity is low. Therefore, it was concluded the assumption of homoscedasticity is met. The results indicate that the overall model significantly predicted STB symptom severity ($F(6, 104)=16.34, p<.001, R^2=.49$). Depressive symptom severity was the only statistically significant predictor variable, such that higher scores on self-reported depression symptoms was associated with higher self-reported STBs (Table 11).

Table 11

Results of Multiple Regression Individual Predictors for STB Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	-.41	1.82		[-4.02, 3.21]	-.22	.82
MFQ	.19	.02	.69	[.15, .23]	9.26	.00**
First-Person Pronoun	.43	.64	.05	[-.86, 1.71]	.66	.51
Sadness	-1.18	1.01	-.09	[-3.19, .82]	-1.17	.25
Anger	-2.01	1.23	-.13	[-4.46, .43]	-1.63	.11
Death	1.12	1.06	.08	[-.98, 3.22]	1.06	.29
Absolutistic	.09	.89	.01	[-1.67, 1.86]	.10	.91

Note. Mood and Feelings Questionnaire (MFQ). ** $p<.001$.

Finally, multiple regression analysis tested the relationship between self-reported anxiety (SCARED) and second/third person pronouns, anxiety, and health word usage (*Hyp 2d*). First, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 9). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.31 or below), indicating that multicollinearity is low

Therefore, it was concluded the assumption of homoscedasticity is met. The results indicate that the overall model significantly predicted anxiety severity ($F(3, 107)=4.38$, $p=.006$, $R^2=.11$). Examination of individual predictors showed that anxiety word usage was the only statistically significant variable, such that higher anxiety word usage was associated with higher self-reported anxiety symptoms (Table 12).

Table 12

Results of Multiple Regression Individual Predictors for Anxiety Symptom Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	1.87	2.59		[-3.25, 6.99]	.73	.47
SPP	-.96	1.42	-.01	[-2.87, 2.76]	-.04	.97
Anxiety	8.81	2.52	.36	[3.18, 13.80]	3.81	.00**
Health	-1.29	1.83	-.07	[-4.94, 2.35]	-.70	.48

Note. Second and third person pronoun (SPP). ** $p < .001$.

Aim 3: Exploratory Analyses

Given that the following research questions do not have hypotheses associated with them, I conducted stepwise regressions to test which domain specific linguistic categories of words would be the most robust in the association with the psychopathology measures.

Research Question 1: Is the severity of STB symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

I conducted a stepwise regression analysis to test the relationship between STB severity and first-person pronoun, second/third-person pronoun, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word usage. First, review of the

histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 10). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.25 or below), indicating that multicollinearity is low. Therefore, it was concluded the assumption of homoscedasticity is met. The final model to predict STB symptom severity included only anxiety word usage ($F(1, 109) = 4.92, p < .05, R^2 = .04$) (Table 13).

Participants who scored higher on self-reported STBs used more anxiety words. None of the other predictors were included in the final model.

Table 13

Results of Stepwise Multiple Regression Individual Predictors for STB Symptom Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	.57	.68		[-.80, 1.91]	.81	.42
Anxiety	3.93	1.78	.21	[.42, 7.43]	2.22	.02*

Note. * $p < 0.05$, ** $p < .01$.

Research Question 2: Is the severity of depressive symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

I conducted a stepwise regression analysis to test the relationship between depression severity and first-person pronoun, second/third person pronoun, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word usage. First, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 11). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.25 or below), indicating that multicollinearity is low. Therefore, it was concluded that

the assumption of homoscedasticity is met. The final model included anxiety and health language variables ($F(2, 108)=7.79, p<.05, R^2=.13$). Participants who scored higher on self-reported depression symptoms used more words related to anxiety and health (Table 14).

Table 14

Results of Stepwise Multiple Regression Individual Predictors for Depression Symptom Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	.16	3.18		[-.50, 9.21]	1.78	.08
Anxiety	14.84	6.98	.21	[1.00, 28.67]	2.13	.03*
Health	10.07	4.98	.20	[.21, 19.94]	2.02	.04*

Note. * $p<0.05$, ** $p<.01$.

Research Question 3: Is the severity of anxiety symptoms associated with greater use of words reflecting self or other focus, negative emotions, and identity domains?

I conducted a stepwise regression analysis to test the relationship between anxiety severity and first-person pronoun, second/third-person pronouns, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word usage. First, review of the histogram, normal PP plot, and scatterplot of the residuals does not indicate a major problem of heteroscedasticity in this model (Appendix A, Figure 12). The variance inflation factor statistic (VIF) indicated that all dependent variables had a tolerable VIF (1.26 or below), indicating that multicollinearity is low. Therefore, it was concluded the assumption of homoscedasticity is met. The final model included anxiety, work, friend, family, and sex language variables (F change (5, 105)=8.60, $p<.05, R^2=.29$). Results

indicated that participants who scored higher on self-reported anxiety symptoms used more words related to anxiety, work, and family (Table 15). However, there was a negative relationship for friend and sex word usage, such that participants who scored higher on self-reported anxiety symptoms used less words associated with friend and sex (Table 15).

Table 15

Results of Stepwise Multiple Regression Individual Predictors for Anxiety Symptom Severity

Variable	B	SE	Beta (β)	95% CI (β)	t	p
Constant	-2.01	1.65		[-5.27, 1.25]	-1.22	.23
Anxiety	8.94	2.07	.36	[4.84, 13.05]	4.32	.00**
Work	3.02	.99	.27	[1.06, 4.99]	3.05	.00*
Friend	-2.59	1.27	-.17	[-5.10, -.89]	-2.04	.00*
Family	3.73	1.11	.31	[1.52, 5.94]	1.52	.00*
Sex	-2.59	1.12	-.17	[-5.33, -.07]	-2.77	.04

Note. * $p < .01$, ** $p < .001$.

CHAPTER IV

DISCUSSION

The purpose of this study was to investigate whether there are detectible differences in naturalistic language use in adolescents with mental illness, with a primary goal of detecting differences amongst adolescents who do, and do not, experience STBs. A secondary goal was to investigate whether inferences could be made, based on natural language use, that would indicate salient identity domains for adolescents with mental illness and with or without STBs. While previous studies have analyzed written language in many forms, including essays, poetry, books, and more recently, discussion forums, to my knowledge, this is the first study that utilizes language data collected from adolescents in a naturalistic setting. Furthermore, this study was one of few that examined linguistic differences in adolescents with mental illness and STBs.

Aim 1

The goal of Aim 1 was to determine whether there were differences in language use between adolescents with psychopathology, adolescents with STBs and no attempts, and adolescents with STBs and at least one prior attempt. Contrary to my hypothesis, there was not a significant main effect of group on first-person pronouns, sadness, anger, death, and absolutistic language use. The only statistically significant difference was observed in pairwise comparisons between the psychiatric control and the suicide attempter groups, such that the attempters used significantly more words related to death than the control group. This indicates that adolescents who have attempted suicide communicate more about death than adolescents for whom suicide is not a concern. This

is consistent with findings from a recent systematic review indicating people who endorse suicidality use significantly more words associated with death (Homan et al., 2022).

One of the predictions of this hypothesis was that there would be differences in first-person pronoun use. This hypothesis was rooted in a rich scientific literature base that has shown a positive relationship between first-person pronoun use and mental illness and suicidality (e.g., Homan et al., 2022). Surprisingly, no differences in first-person pronoun use were detected. It could be that because the adolescents in this sample all had mental illness and were not compared to adolescents without mental illness, suggesting that experiencing a mental illness in and of itself might lend itself to more self-introspection, and therefore more usage of first-person pronouns. Indeed, theories attempting to explain the relationship between first-person pronoun use and mental illness have suggested as much (Watkins & Teasdale, 2004). However, there remains a lack of consistency in this relationship. For example, a recent investigation of language use amongst adolescents at risk for psychosis also did not find differences in first-person pronoun use between those who did or did not endorse suicidality (Dobbs et al., 2023).

My investigation also did not reveal any group differences in sadness and anger words. Homan and colleagues (2022) reported that many studies have indicated that words related to negative emotionality, including sadness and anger words, are associated with suicidality, with a few studies that showed that people who engaged in suicidal behaviors used these word categories more than people who endorsed suicidal thoughts without suicidal behaviors. In Dobbs and colleagues' (2023) investigation, they found that suicidality was positively associated with anger words, but not sadness words. The inconsistency in these relationships is found across the literature. One of the reasons for

the variation could be due to the heterogenous nature of language data. In their systematic review, Homan and colleagues (2022) reported that language data were collected from a variety of sources, including social media sites, medical records, therapy notes, suicide notes, text message, and voice samples. It could be that there is contextual variation in how people use language. Therefore, the assumption that people use language similarly on a public social media post, for example, as well as private text messages, may be inaccurate.

Additionally, studies varied on how they operationalized suicidality (Homan et al., 2022). Definitions of suicidality included review of medical records and therapy notes in which a health professional documented suicidality, self-report suicidality questionnaires, and self-disclosure on social media platforms. Furthermore, many studies did not differentiate between suicidal thoughts and suicidal behaviors. The diversity in the nature of language use data and measures of suicidality could provide one explanation for the heterogenous nature of findings.

My hypothesis that absolutistic words would be used more by adolescents who endorsed suicidality was not supported. This hypothesis stemmed from more recent work suggesting that use of absolutistic words (e.g., always, never, forever) are indicative of a sense of finality or extremes and are associated with suicidality (Adam-Troian & Arciszewski, 2020; Al-Mosaiwi & Johnstone, 2018). Certainly, the idea of death by suicide as the only option to end emotional pain would be considered extreme and final. Al-Mosaiwi and Johnstone (2018) found that absolutistic words were used more on depression, anxiety, and suicidal ideation discussion forums, compared to control. They also found that absolutistic words were used most on suicidal ideation forums. While this

suggests there may be specificity in regard to absolutistic word use and suicidality, the use of more absolutistic words was also present in the depression and anxiety forums. A potential limitation of these findings is that there were no diagnostic or symptom data to describe the forum writers in the sample. Therefore, it is not possible to confirm the writers' mental illnesses, nor whether they were experiencing symptoms of depression, anxiety, or suicidality at the time. In Adam-Troian and Arciszewski's (2020) study, they found a predictive positive relationship between use of absolutistic words in Google search queries and occurrence of death by suicide in the United States. While these findings show a relationship between absolutistic word use and suicide rates, there is no way to discern whether it was the people experiencing suicidal thought and behaviors who used more absolutistic words.

In attempt to further isolate language differences attributable to suicidality specifically, I included depressive symptoms in the next model. There were no group differences in language use between psychiatric control, suicide ideators, and suicide attempters when including depressive symptoms in the model. Depressive symptoms were positively associated with words related to sadness and anger. This finding is consistent with the research on the relationship between negative emotion language, including sadness and anger words, and depressive symptoms. A recent meta-analysis of 13 studies found a medium effect size (Cohen's $d = .72$) for the relationship between depressive symptoms and negative emotion language use (Tølbøll, 2019). (Rude, Gortner, & Pennebaker, 2001). The positive relationship between anger language use and depressive symptoms is consistent with conceptualization of adolescent depression, such

that anger, or irritability, is often a more common symptom of depression during this period of life than in adulthood (American Psychiatric Association, 2022).

Aim 2

The goal of Aim 2 was to test the hypotheses that there are positive relationships between specific language use categories (pronouns, sadness, anger, anxiety, health, death, and absolutist words) and quantitative measures of STBs (*Hyp2a/c*), depressive symptoms (*Hyp2b*), and anxiety symptoms (*Hyp2d*). Previous research has provided evidence that there are differences in language use in people with mental illness and suggests there may be specificity in regard to language use and type of symptoms (Adam-Troian & Arciszewski, 2020; Al-Mosaiwi & Johnstone, 2018; Coppersmith et al., 2015; Nobles et al., 2018). I predicted that the severity of self-reported STBs and depressive symptoms would be associated with greater use of first-person pronouns, sadness, anger, death, and absolutist words. STBs and depressive symptoms were both positively associated with words in the death category, but none of the other language use categories were significant predictors. Furthermore, when including STBs and depressive symptoms in the same model, depressive symptoms alone accounted for a significant proportion of the variance in death word category. This was counter to my hypothesis that STB symptoms would be a unique predictor, above and beyond depressive symptoms. This finding is not wholly unexpected given that STBs are most often comorbid with experience of depression.

It was surprising that first-person pronouns, anger, and sadness were not significant predictors, as these have been some of the more consistent language use variables associated with depression and suicidality (e.g., Coppersmith et al., 2015;

Lyons et al., 2018; Rude et al., 2004). Additionally, Tølbøll's (2019) meta-analysis found a small to medium effect (Cohen's $d = .44$) for the relationship between depressive symptoms and first-person pronoun use and a medium effect (Cohen's $d = .72$) for the relationship between depressive symptoms and negative emotion language use. One explanation for this discrepancy could lie in the type of language use data used in much previous work and language use data used in this study. Tølbøll's (2019) meta-analysis reported that the studies used a variety of language data including written language, in the form of essays, blog posts, Facebook posts and transcribed oral language in the form of semi-structured interviews, interviews with therapists, free-response to photographs, and "audio-recordings from the everyday lives of depressed people (p. 49)." It could be that there are contextual differences in language use intended for public consumption and language use intended to be kept private between individuals. Additionally, researchers investigated whether depressive symptoms were associated with negative emotion language use on Facebook and Twitter posts (Seabrook et al., 2018). They found that negative emotion language use on Facebook was positively associated with depressive symptoms, however, negative emotion language use on Twitter was negatively associated with depressive symptoms. This finding provides support for the hypothesis that language use is context dependent.

I predicted that the severity of self-reported anxiety symptoms would be positively associated with words associated with second- and third-person pronouns (e.g., you, she, he, they), anxiety, and health language categories. Anxiety words were the only language use category associated with anxiety symptoms. This finding is encouraging in that it makes conceptual sense for anxiety word use to be associated with symptoms of

anxiety. My prediction that words related to health would be associated with anxiety symptoms was not supported. Previous research has indicated that worries about health are common amongst children and adolescents with anxiety disorders (Weems, Silverman & La Greca, 2000). While worries about health may be a common concern for youth with anxiety disorders, it is not a ubiquitous symptom of anxiety.

In addition, my hypothesis that anxiety symptoms would be associated with greater use of second- and third-person pronouns was not supported. This hypothesis was generated based on research that has indicated that anxiety symptomology is associated with higher use of second- and third-person pronouns (Sonnenschein, Hofmann, Ziegelmayer, & Lutz, 2018). The use of second- and third- person pronouns may indicate a greater focus on others, which conceptually could represent a symptom of social anxiety in which a person is concerned with by negative judgements of others. However, similar to health concerns, worries about the judgement of others is not a ubiquitous symptom of anxiety disorders, in general. Research investigating the linguistic differences between anxiety disorders found that people with generalized anxiety used more words related to social processes and people with social anxiety used more anger, sexual, and swear words (Stamatis et al., 2022). This finding suggests that the differences in anxiety disorders, and their associated symptoms, are reflected in language use as well.

Aim 3

In Aim 3, I investigated whether patterns of online natural language use can detect identity distress among adolescents experiencing psychopathology. While this study does not include a previously validated measure of identity formation, nor identity distress, I suggested that some language use categories may be representative of important domains

in identity formation. Domains shown to be important in adolescent identity development are represented as language use categories of friend, family, work, religion, health, and sex (Goossens, 2001; van Doeselaar et al. 2018). I theorized that domain specificity may provide insight into areas of importance for adolescents who experience STBs, depression, and anxiety. Given the exploratory nature of this aim, I used stepwise regressions in order to find the best fitting models.

For Research Question 1, I asked whether first-person pronoun, second/third person pronoun, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word usage to find the best fitting model to predict STB symptoms. The best fitting model included only anxiety word usage. None of the other language use variables were significant predictors. It was somewhat surprising that anxiety language use was the only variable that accounted for variance in STB symptoms. However, some research has indicated that anxiety symptoms often co-occur in people with suicidality. In a sample of 2,778 adults receiving outpatient psychiatric care, researchers found that people with moderate anxiety symptoms were two times as likely to report suicidality, even when controlling for depressive symptoms (Diefenbach, Woolley, & Goethe, 2009). Another study investigated the relationship between anxiety disorders and the prevalence of suicidality (n=4,131 adults, data from National Co-morbidity Survey-Replication) and found panic disorder, social anxiety disorder (SAD), posttraumatic stress disorder (PTSD), and generalized anxiety disorder (GAD) diagnoses were positively related to suicidal ideation (Cogle, Keough, Riccardi & Sachs-Ericsson, 2009). Furthermore, Cogle and colleagues (2009) found that SAD, PTSD, and GAD were associated with a higher likelihood of suicide attempts.

More recently, researchers used network analysis to investigate the relationship between anxiety and depressive symptoms in adolescents with suicidal ideation and found that endorsement of worrying too much, guilt, and irritability were the primary symptoms linked to suicidality (Cai et al., 2023). This highlights the need to investigate the role anxiety plays in suicidality, in addition to the more common assessment of depressive symptoms. Results from the current study suggest that increases in anxiety language use may be a potential marker for increased suicidality.

For research question 2, I asked whether first-person pronoun, second/third-person pronoun, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word usage to find the best fitting model to predict depressive symptom severity. The best fitting model showed that anxiety words and health words accounted for the most variance in depressive symptoms. The relationship between anxiety language use and depressive symptoms may be reflective of the frequent co-morbidity of depression and anxiety. In addition, this points to the importance of considering anxiety language use in the presence of depression, similarly to the relationship between anxiety language use and suicidality, previously discussed. The final model also showed that health language use was positively associated with depression, suggesting that adolescents who experience more depressive symptoms communicate more about health. Again, it was surprising that sadness and anger language use variables were not included in the final model, as these negative emotion language categories have been replicated in the natural language use literature (Rude et al., 2004; Seabrook, Kern, Fulcher & Rickard, 2018).

For research question 3, I analyzed first-person pronoun, second/third-person pronouns, sadness, anger, anxiety, friend, family, work, religion, health, and sexual word

usage to find the best fitting model to predict anxiety symptoms severity. The best fitting model included anxiety, work, friends, family, and sex words, and showed a positive relationship with anxiety symptoms. This finding makes sense from a conceptual understanding of anxiety, and particularly generalized anxiety, in which people often have worries in multiple domains (American Psychiatric Association, 2022).

Taken together, the domains of work, friends, family, health, and sex language use categories showed some relationship to depressive and anxiety symptoms. However, none of the hypothesized identity domain language use categories were significantly related to suicidality. Reflecting back, I have considered whether it would have been better to exclude language use categories related to symptomatology (e.g., sadness, anger, and anxiety); only including words related to the hypothesized identity domains in order to isolate potentially salient domains.

Limitations and Strengths

While this study provides some evidence of meaningful linguistic differences in adolescents with mental illness and suicidality, some limitations should be taken into account. Firstly, small differences in the average proportions of linguistic categories may have made it difficult to detect differences in language use with regression analyses used in this study. Analysis strategies that use overall mean differences are not able to detect potential differences in language use that could vary as function of transient mood states. For example, the experience of suicidality is not often considered a continuous state of mind, rather it fluctuates in its presence and intensity (Kleiman et al., 2017). Therefore, using mean level difference analysis of language use over the course of months is not

able to isolate what may be a short period of suicidality, from the overall language use which likely included much more time of non-suicidality.

Second, even though the sample size is fairly large for a high-risk sample such as this, a larger sample size may provide more power and may be more robust against Type II errors. The use of between-group analyses for some hypotheses further reduced the power in this study.

Thirdly, the use of stepwise regression (used in Aim 3) is generally cautioned because of the increased risk for Type I errors. While the use of stepwise regression is considered appropriate for analyses in which there is no specific hypotheses, results from these analyses should be interpreted with caution.

Additionally, there was no validated measure of identity formation states or processes in Aim 3. I made inferences about important identity domains based on language use, but there is no research to date to support this inference. The inclusion of emotion word use categories, particularly anxiety word use, in these analyses could also have affected the detectability of differences in language use by domain, as these are more robustly associated with psychopathology, and could be accounting for most of the variance. It could also be that all of these domains represent significant areas for most adolescents, and it may therefore be difficult to detect subtle differences. Conversely, there may be individual differences in what domains are considered important for a given adolescent. Also, the inference that more language use for certain identity domains would be associated with increased symptoms of mental illness fails to consider that focus on a particular domain may be experienced in a positive or negative fashion. Identity formation processes and increase salience of these domains are considered a normative

aspect of adolescent development and an adolescent's increased focus on their role in peer relationships, for example, could be helpful or harmful, depending on the context (Goossens, 2001; van Doeselaar et al. 2018). In addition, identity formation is considered a dynamic process and mean-level, or between-group, differences does not capture variance or changes on an individual level (Schwartz et al., 2011).

In addition, the relationship between age and identity formation states and processes are significant when considering mental illness. A given identity state (e.g., commitment) or identity process (e.g., in-depth exploration) is not indicative of mental well-being or illness in and of themselves. The age at which these processes occur is related to the likelihood of mental health or illness. For example, younger adolescents are more likely to be in uncommitted identity states (e.g., moratorium) which is not associated with increase in mental illness (Luyckx et al., 2013). However, when older adolescents are in uncommitted identity states, the association with mental illness strengthens (Crocetti, et al., 2009). While identity formation is dynamic in nature, there seems to be consistency around when identity crises become problematic, particularly in regards to age.

While there are some limitations to this study, there are significant strengths as well. This study used passively collected out-going text-messaging communication from adolescents' digital devices over the course of six months. Much language use research in the digital space has relied on public posts made on social media platforms and online discussion forums (Homan et al., 2022; Nanomi Arachchige, Sandanapitchai & Weerasinghe, 2021; Young et al., 2023). While there is certainly an abundance of important questions to investigate in that space, there may be differences in how people

use language when making a public post versus sending private messages. Given this important distinction, the naturalistic data used in this study are particularly important for adolescents because they spend a significant amount of time using digital devices for communication, and this is frequently their preferred method of communication (Lenhart et al., 2015).

Additionally, this study is one of few that that investigated the natural language use of adolescents with psychopathology and suicidality. Much of the research on natural language use and psychopathology has focused on adult samples (Homan et al., 2022). The results from this study suggest there may be differences in how adolescents use language, compared to adults, given that some of the more robust relationships between language use and mental illness were not replicated (e.g., first-person pronoun use, negative emotionality). Findings from this study also indicated there may be specificity in language use for adolescents that varies as a function of suicidality, depression, and anxiety symptomology.

Future Directions

Future research can build upon these findings through replication and by addressing limitations. The nature of these data could allow for hourly, daily, or weekly analysis of language use and analyze changes over time. Therefore, researchers could test the hypothesis that language use will vary as a function of mood states. Certainly, previous work has shown that people who are experiencing psychopathology use language differently (Coppersmith et al., 2015; Dutta & De Choudhury, 2020; Rude et al., 2004; Shen & Rudzicz, 2017). It stands to reason that a given individual's language use would change when they are in a depressed versus non-depressed state, or when they

are experiencing periods of suicidal ideation, for example. Detecting fluctuations in mood states in real time would set the groundwork for the development of just-in-time adaptive interventions that could provide help for adolescents at their precise moment of need (Nahum-Shani et al., 2018).

Additionally, there may be important diversity factors that influence how people use language that warrant investigation. Linguistic anthropologists have recognized that there are differences in how people use language that varies as a function of culture, broadly speaking (Bucholtz & Hall, 2004). More recent work has described methodological strategies for analyzing cultural differences in natural language use that could be applied to language use analysis within the mental health field (Berger & Packard, 2022). The investigation of cultural differences in regard to natural language use and mental illness are particularly important when findings are used to identify risk states and develop interventions.

The language use results from this study could be used as potential starting points for these investigations. For example, the relationship between the use of words related to death and suicidality symptomology is particularly promising, given the conceptual link between the two. The relationship between depressive symptomology and sadness and anger language use also warrants further investigation. In addition, the exploratory analyses revealed that anxiety language use was consistently positively related to anxiety, depressive, and suicidality symptomology. Future research could investigate whether these language use categories are associated with changes in severity of mental illness within an individual.

In this study, I used the average proportion of word per language use category in the analyses. While this is a commonly used analysis in studies of natural language use, this is but one methodology available in language analysis. Other types of language analysis, such as content analyses, attempt to infer context by linking words to phrases, and could be used to determine particular themes that are relevant to mental illness and suicidal ideation. Analyses of linguistic phrases may be able to capture things such as suicidal thoughts and behaviors that might otherwise go missed in analysis of single word use. For example, one study investigated certain phrases in tweets, such as “tired of living” and “I want to disappear” that are suggestive of suicidal thinking (O’dea, Wan, Batterham, Calear, Paris & Christensen, 2015). In single word language analysis, the words tired, living, and disappear would not be included in categories suggestive of suicidal thoughts or behaviors, nor even negative emotionality.

In regard to questions about identity formation, future research could investigate whether there are specific language markers that map onto measures of identity formation states, processes, or domains. Given that identity formation is such a significant part of adolescence, determining areas of struggle in regard to identity would allow for tailoring of interventions. Identification of salient identity domains could help providers target a specific area of weakness, or capitalize on an area of strength, for a given individual.

Conclusion

In sum, this research provides an important starting point for understanding the relationship between on device language use in a natural setting amongst adolescent with mental illness and suicidality. Suicide is one of the leading causes of death in adolescents (Ruch et al., 2019), therefore research that investigates potential markers of increased risk

is imperative. Additionally, adolescent mental illness, in general, not only causes suffering, but is known to increase risk for mental illness later in life, poorer physical health, and poorer social and economic outcomes (Ormel et al., 2017).

This study first reviewed the literature on identity formation in adolescence. Identity formation is the key developmental task of adolescence and more difficulty navigating identity changes is related to adolescent struggle of mental illness (Berman, 2019; Meeus, 2011). Next, this study reviewed the literature on the relationship between natural language use and mental illness and suicidality. This review identified language use categories associated with mental illness and suicidality in primarily adult samples, including pronoun use, emotionality, absolutism, and death word use. Next, I suggested the importance of investigation of natural language use in relation to identity formation in adolescents struggling with mental illness. I made predictions that natural language use findings in adult samples would be replicated in a high-risk adolescent sample. I then proposed an exploratory investigation into whether natural language use categories that were conceptually related to important identity domains would be differentially related to internalizing symptomology and suicidality.

This study found that many of the word categories shown to be consistent in investigations of natural language use in adults with mental illness (e.g., first-person pronouns and negative emotionality), were not replicated in this high risk adolescent sample. I discussed potential explanations for the lack of replication and provided suggestions for future investigations. I highlighted the conceptual significance of the finding that adolescent suicidality was positively related to the use of death words and

suggested further investigation of whether death word use could be a marker of increase in suicide risk.

While this study's hypotheses about the relationship between language use and identity domains were not statistically robust, this remains an important and novel area of exploration. I provided suggestions, including the use of previously validated measures of identity formation and individual analysis, as a strategy to further this investigation.

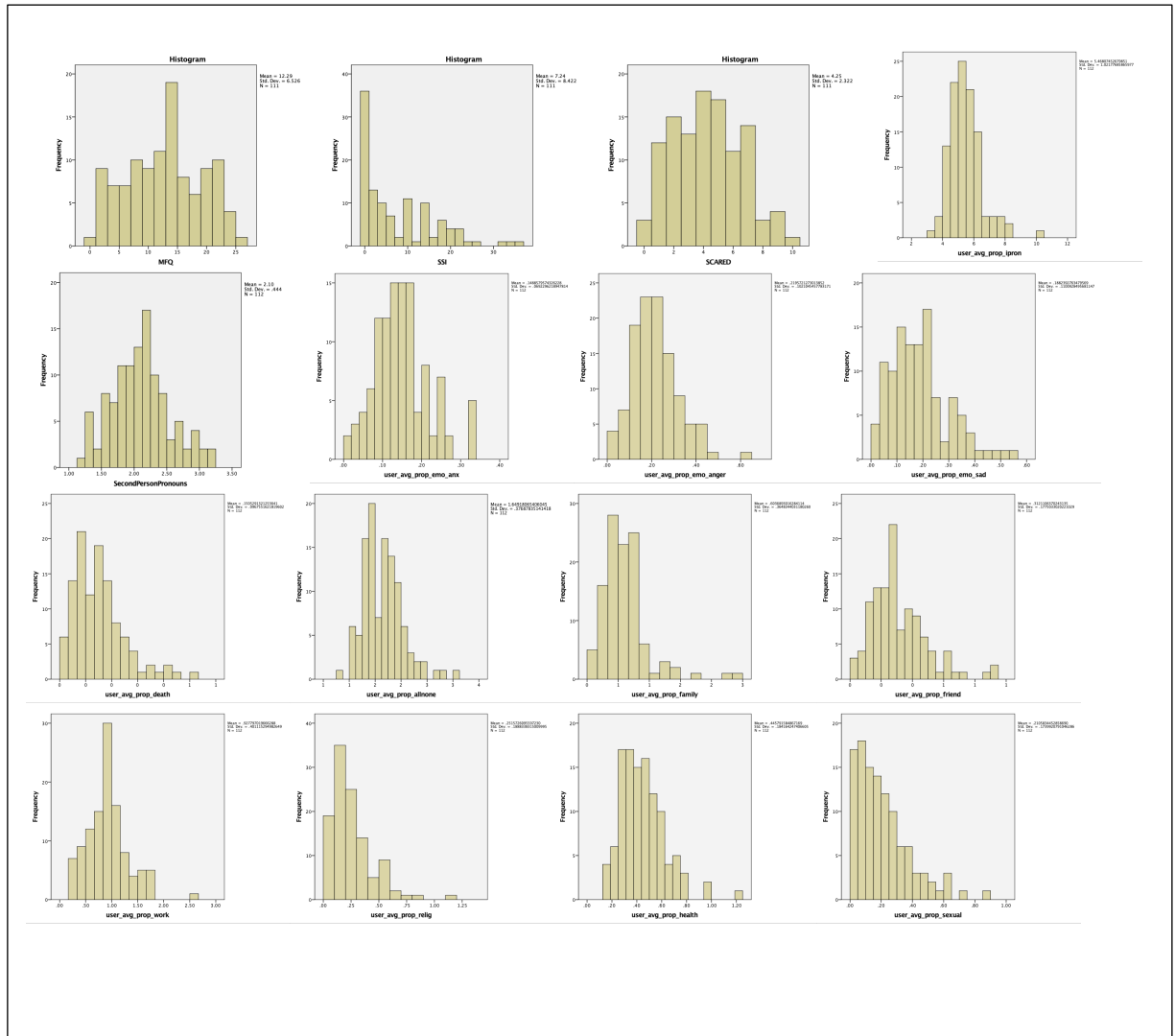
In conclusion, language use analysis is a meaningful and powerful tool that has the potential to capture changes affective states in real time, creating the opportunity for quick and specific intervention.

APPENDIX A

Figures

Figure 1

Histograms of all Variables Prior to Data Transformation



Note. Cognitive processing language category (user_avg_cogproc_prop), negative emotion language category (user_avg_neg_prop), and first person pronoun language category (user_avg_ipron_prop).

Figure 2

Box and Whisker Plots of all Variables Prior to Data Transformation

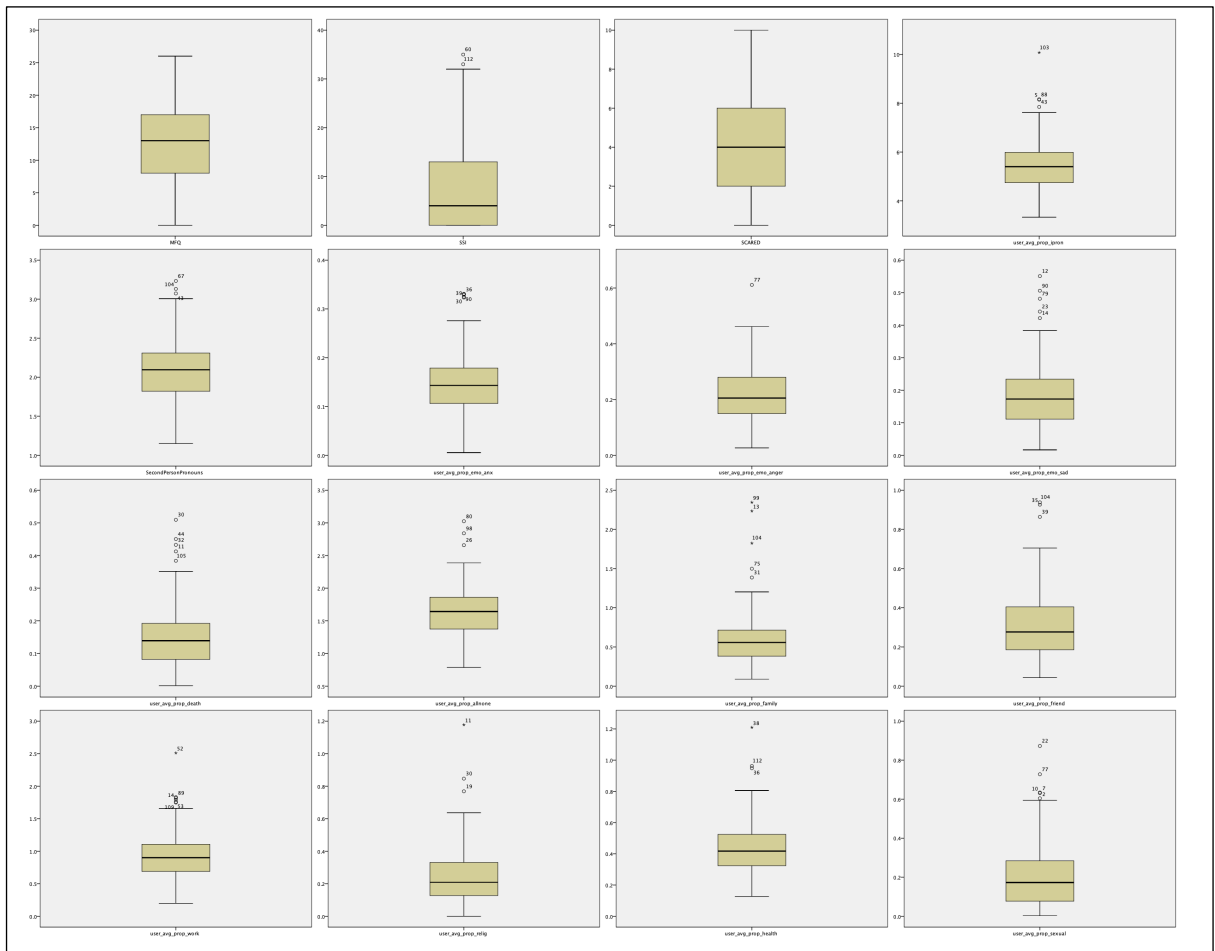


Figure 3

QQ Plots of all Variables Prior to Data Transformation

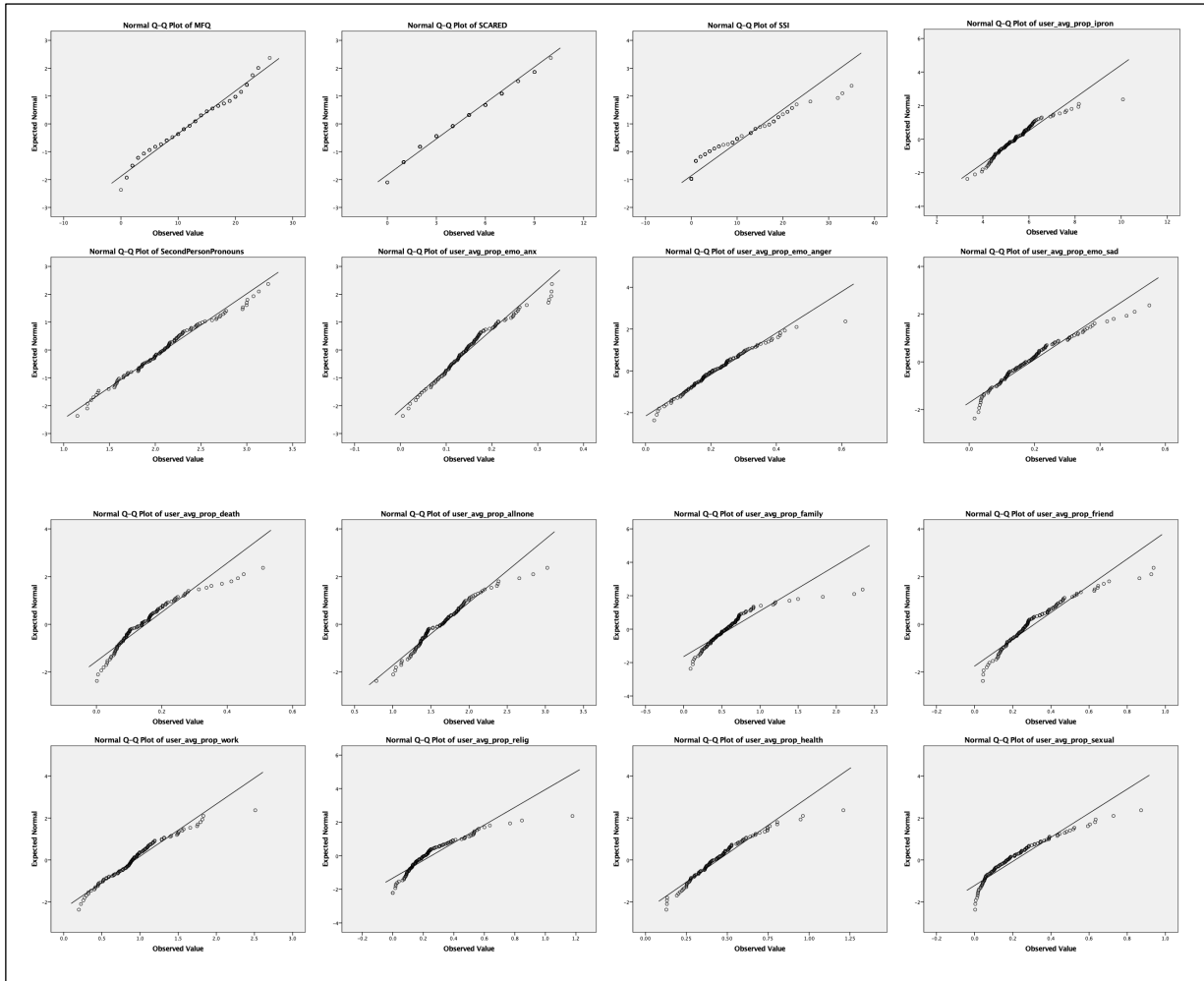


Figure 4

Histograms of Square Root Transformed Variables

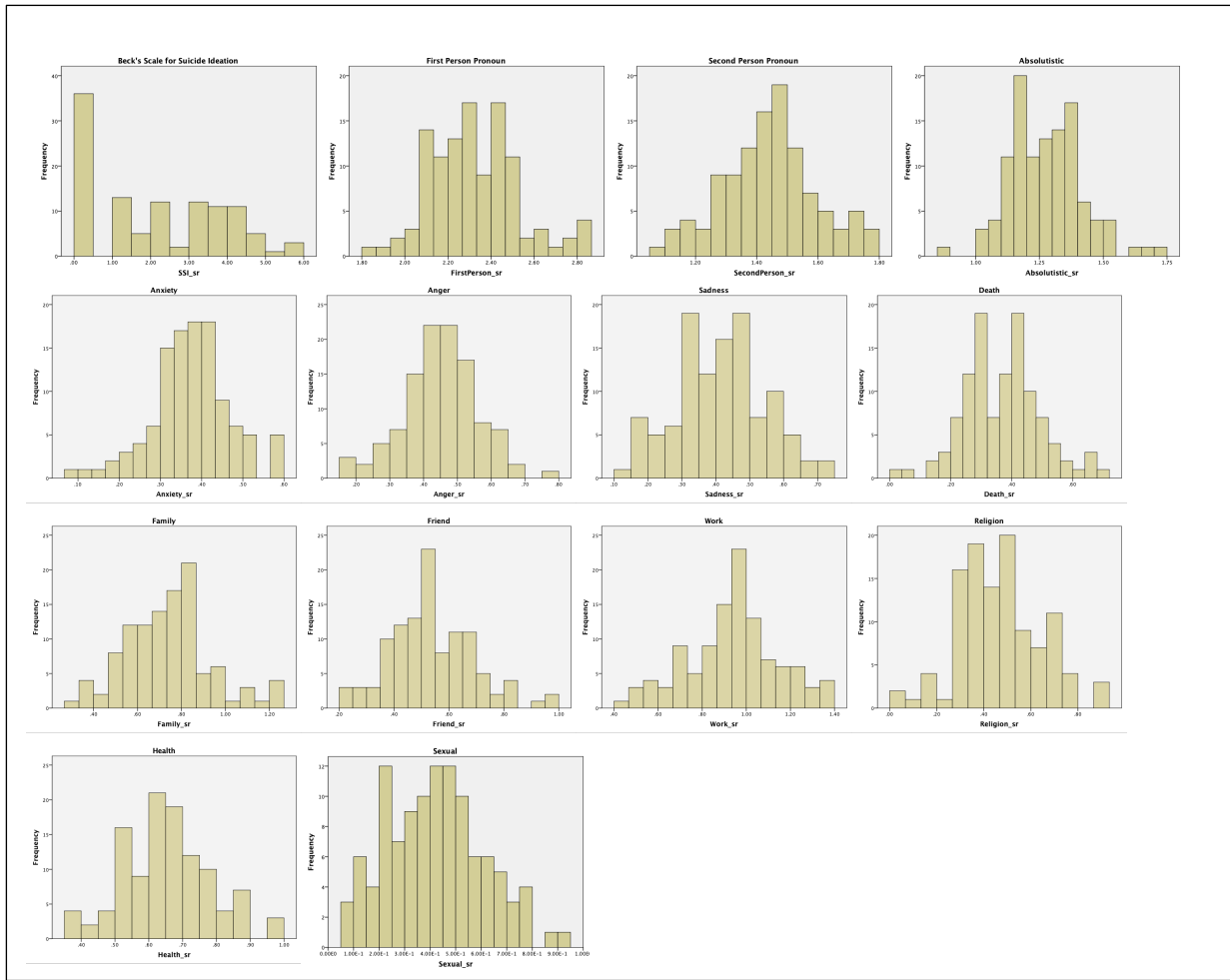


Figure 5

Normal QQ Plots of Square Root Transformed Variables

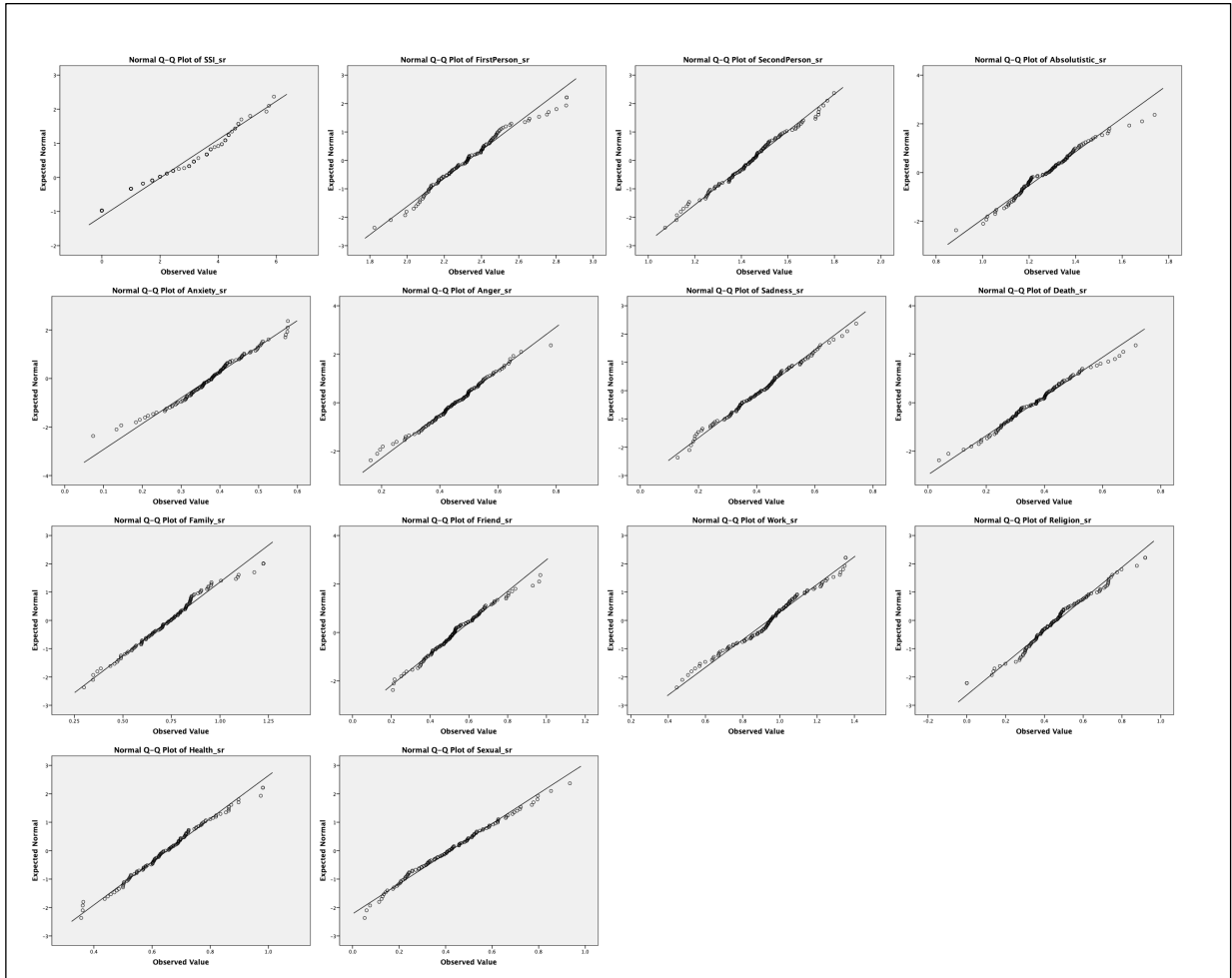


Figure 6

Histogram, Normal PP Plot, Scatterplot of Residuals Aim 2 (Hyp2a)

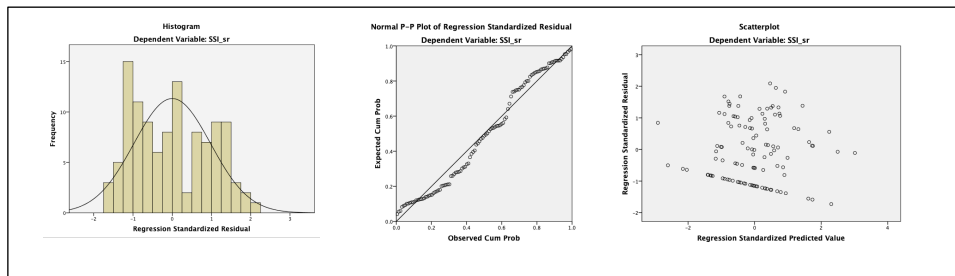


Figure 7

Histogram, Normal PP Plot, Scatterplot of Residuals Aim 2 (Hyp2b)

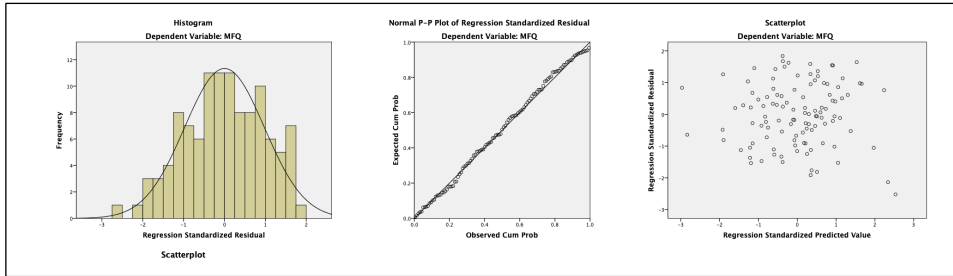


Figure 8

Histogram, Normal PP Plot, Scatterplot of Residuals Aim 2 (Hyp2c)

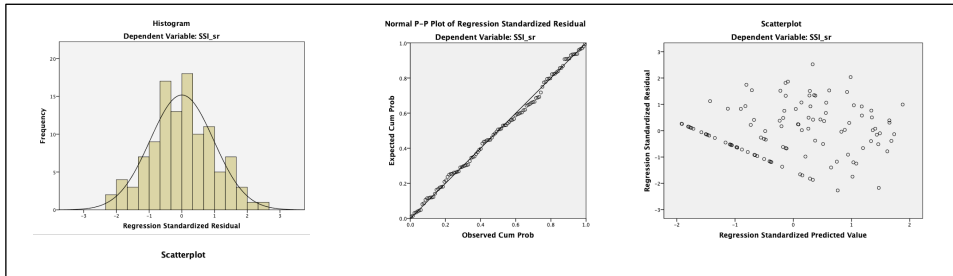


Figure 9

Histogram, Normal PP Plot, Scatterplot of Residuals Aim 2 (Hyp2d)

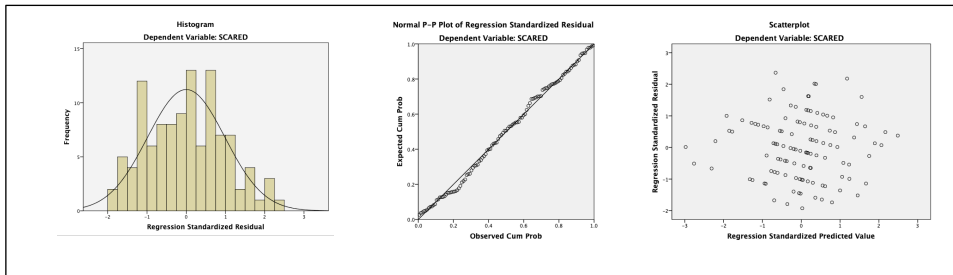


Figure 10

Histogram, Normal PP Plot, Scatterplot of Residuals Exploratory Question 1

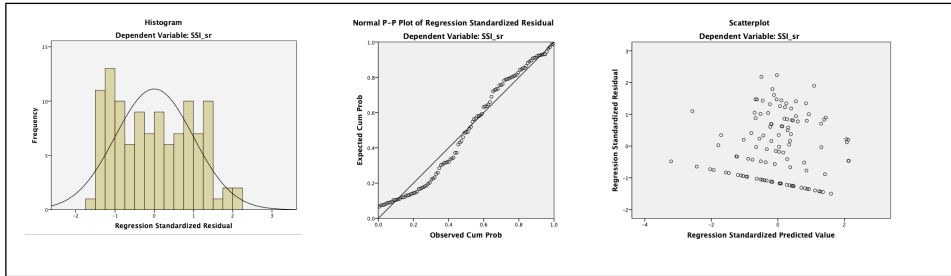


Figure 11

Histogram, Normal PP Plot, Scatterplot of Residuals Exploratory Question 2

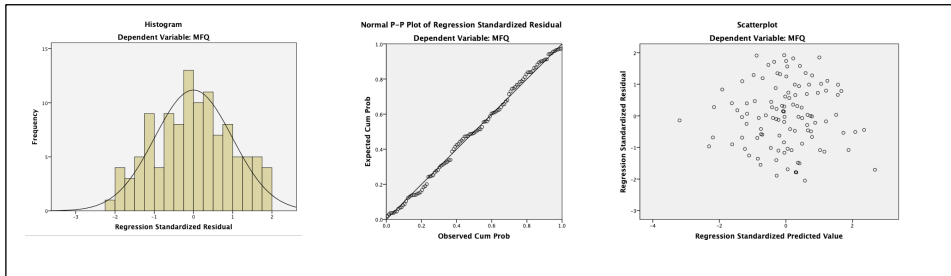
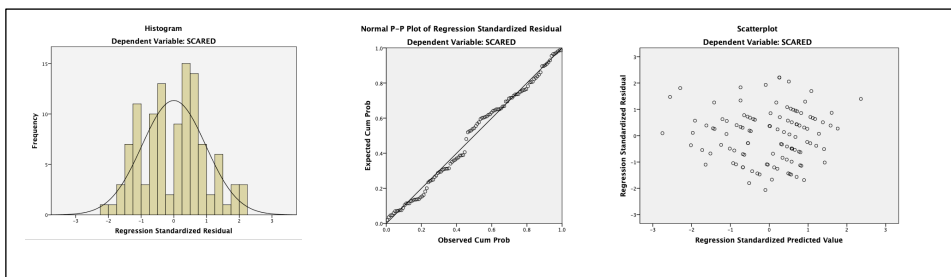


Figure 12

Histogram, Normal PP Plot, Scatterplot of Residuals Exploratory Question 3



APPENDIX B

TABLES

Table 1

Lilliefors corrected Kolmogorov-Smirnov (K-S) and Shapiro-Wilk Tests After Transformation

<u>Self report variable</u>	<u>K-S (p)</u>	<u>Shapiro-Wilk (p)</u>
SSI	.198*	.892*
<u>Language variable</u>		
FPP	.065	.976*
SPP	.058	.988
Anxiety	.098	.981
Anger	.047	.992
Sad	.053	.991
Death	.062	.989
Absolutistic	.098*	.981
Family	.088*	.977*
Friend	.088*	.983
Work	.085*	.918
Religion	.088*	.983
Health	.062	.989
Sex	.051	.988

Note. SSI = Beck's Scale for Suicide Ideation, FPP = first person pronouns, SPP = second person pronouns. * $p < .05$.

APPENDIX C

Self-report Questionnaires

BECK'S SCALE FOR SUICIDE IDEATION (Beck, Kovacs & Weissman, 1979)

1. Wish to live.
2. Wish to die.
3. Reasons for living/dying.
4. Desire to make active suicide attempt
5. Passive suicidal desire.
6. Duration of suicidal thoughts
7. Frequency of ideation
8. Attitude toward ideation
9. Control over suicidal ideation
10. Deterrents to attempt
11. Reasons for attempt
12. Specificity of planning
13. Availability or opportunity
14. Capability to carry out attempt
15. Expectancy of actual attempt
16. Extent of actual preparation
17. Suicide note
18. Final acts
19. Deception and concealment

APPENDIX D

MOOD AND FEELINGS QUESTIONNAIRE (Angold et al., 1995)

1. I felt miserable or unhappy.
2. I didn't enjoy anything at all.
3. I felt so tired I just sat around and did nothing.
4. I was very restless.
5. I felt I was no good anymore.
6. I cried a lot.
7. I found it hard to think properly or concentrate.
8. I hated myself.
9. I was a bad person.
10. I felt lonely.
11. I thought nobody really loved me.
12. I thought I could never be as good as other kids.
13. I did everything wrong.

APPENDIX E

SCREEN FOR ANXIETY RELATED EMOTIONAL DISORDERS (Birmaher et

al., 1999)

1. I get really frightened for no reason at all.
2. I am afraid to be alone in the house.
3. People tell me that I worry too much.
4. I am scared to go to school.
5. I am shy.

REFERENCES CITED

- Albarello, F., Crocetti, E., & Rubini, M. (2018). I and us: A longitudinal study on the interplay of personal and social identity in adolescence. *Journal of youth and adolescence, 47*(4), 689-702.
- 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.
- American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). <https://doi.org/10.1176/appi.books.9780890425787>
- Bar-Joseph, H., & Tzuriel, D. (1990). Suicidal tendencies and ego identity in adolescence. *Adolescence, 25*(97), 215.
- Becht, A. I., Nelemans, S. A., Branje, S. J., Vollebergh, W. A., Koot, H. M., & Meeus, W. H. (2017). Identity uncertainty and commitment making across adolescence: Five-year within-person associations using daily identity reports. *Developmental Psychology, 53*(11), 2103.
- Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: the Scale for Suicide Ideation. *Journal of consulting and clinical psychology, 47*(2), 343.
- Bell, B. T. (2019). “You take fifty photos, delete forty nine and use one”: A qualitative study of adolescent image-sharing practices on social media. *International Journal of Child-Computer Interaction, 20*, 64-71.
- Berger, J., & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist, 77*(4), 525.
- Berman, S. L. (2019). Identity distress. *The Encyclopedia of Child and Adolescent Development, 1-11*.
- Beyers, W., & Luyckx, K. (2016). Ruminative exploration and reconsideration of commitment as risk factors for suboptimal identity development in adolescence and emerging adulthood. *Journal of adolescence, 47*, 169-178.
- Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The Development and Psychometric Properties of LIWC-22.
- Brent, D. A., Johnson, B. A., Perper, J., Connolly, J., Bridge, E. J., Bartle, S., & Rather, E. (1994). Personality disorder, personality traits, impulsive violence, and completed suicide in adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*, 1080-1086.

- Bucholtz, M., & Hall, K. (2004). Language and identity. A companion to linguistic anthropology, 1, 369-394.
- Bucholtz, M., & Hall, K. (2010). Locating identity in language. *Language and identities*, 18, 28.
- Cai, H., Chow, I. H., Lei, S. M., Lok, G. K., Su, Z., Cheung, T., ... & Xiang, Y. T. (2023). Inter-relationships of depressive and anxiety symptoms with suicidality among adolescents: A network perspective. *Journal of Affective Disorders*, 324, 480-488.
- Chandler, M. J., Lalonde, C. E., Sokol, B. W., Hallett, D., & Marcia, J. E. (2003). Personal persistence, identity development, and suicide: A study of native and non-native North American adolescents. *Monographs of the society for research in child development*, i-138.
- Coppersmith, G., Leary, R., Whyne, E., & Wood, T. (2015). Quantifying suicidal ideation via language usage on social media. In *Joint Statistics Meetings Proceedings, Statistical Computing Section*, JSM.
- Cogle, J. R., Keough, M. E., Riccardi, C. J., & Sachs-Ericsson, N. (2009). Anxiety disorders and suicidality in the National Comorbidity Survey-Replication. *Journal of psychiatric research*, 43(9), 825-829.
- Crocetti, E. (2017). Identity formation in adolescence: The dynamic of forming and consolidating identity commitments. *Child Development Perspectives*, 11(2), 145-150.
- Crocetti, E., Klimstra, T., Keijsers, L., Hale, W. W., & Meeus, W. (2009). Anxiety trajectories and identity development in adolescence: A five-wave longitudinal study. *Journal of Youth and Adolescence*, 38(6), 839-849.
- Crocetti, E., Rubini, M., & Meeus, W. (2008). Capturing the dynamics of identity formation in various ethnic groups: Development and validation of a three-dimensional model. *Journal of adolescence*, 31(2), 207-222.
- Cyr, B. A., Berman, S. L., & Smith, M. L. (2015). The role of communication technology in adolescent relationships and identity development. In *Child & Youth Care Forum* 44(1) 79-92.
- De Choudhury, M., Sharma, S. S., Logar, T., Eekhout, W., & Nielsen, R. C. (2017, February). Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 353-369).

- Diefenbach, G. J., Woolley, S. B., & Goethe, J. W. (2009). The association between self-reported anxiety symptoms and suicidality. *The Journal of nervous and mental disease, 197*(2), 92-97.
- Dixon, W. J. (1980). Efficient analysis of experimental observations. *Annual review of pharmacology and toxicology, 20*(1), 441-462.
- Dobbs, M. F., McGowan, A., Selloni, A., Bilgrami, Z., Sarac, C., Cotter, M., ... & Srivastava, A. (2023). Linguistic correlates of suicidal ideation in youth at clinical high-risk for psychosis. *Schizophrenia Research*.
- Dutta, S., & De Choudhury, M. (2020). Characterizing Anxiety Disorders with Online Social and Interactional Networks. In *International Conference on Human-Computer Interaction* (pp. 249-264). Springer, Cham.
- Erikson, E. H. (1968). Identity: Youth and crisis (No. 7). WW Norton & company.
- Foto-Ozdemir, D., Akdemir, D., & Cuhadaroglu-Cetin, F. (2016). Gender differences in defense mechanisms, ways of coping with stress and sense of identity in adolescent suicide attempts.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism, 10*(2), 486.
- Goossens, L. (2001). Global versus domain-specific statuses in identity research: A comparison of two self-report measures. *Journal of adolescence, 24*(6), 681-699.
- Gray, L. (2018). Exploring how and why young people use social networking sites. *Educational Psychology in Practice, 34*(2), 175-194.
- Guz, S., Kattari, S. K., Atteberry-Ash, B., Klemmer, C. L., Call, J., & Kattari, L. (2021). Depression and suicide risk at the cross-section of sexual orientation and gender identity for youth. *Journal of Adolescent Health, 68*(2), 317-323.
- Hale III, W. W., Raaijmakers, Q., Muris, P., & Meeus, W. (2008). Developmental trajectories of adolescent anxiety disorder symptoms: A 5-year prospective community study. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(5), 556-564.
- Hall, G. S. (1905). *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion and education* (Vol. 2). D. Appleton.
- Hancock, J. T., & Dunham, P. J. (2001). Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions. *Communication research, 28*(3), 325-347.

- Hernandez, L., Montgomery, M. J., & Kurtines, W. M. (2006). Identity distress and adjustment problems in at-risk adolescents. *Identity, 6*(1), 27-33.
- Hoekstra, R., Kiers, H. A., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology, 3*, 137.
- Holi, M. M., Pelkonen, M., Karlsson, L., Kiviruusu, O., Ruutu, T., Heilä, H., ... & Marttunen, M. (2005). Psychometric properties and clinical utility of the Scale for Suicidal Ideation (SSI) in adolescents. *BMC psychiatry, 5*(1), 1-8.
- Homan, S., Gabi, M., Klee, N., Bachmann, S., Moser, A. M., Michel, S., ... & Kleim, B. (2022). Linguistic features of suicidal thoughts and behaviors: A systematic review. *Clinical psychology review, 95*, 102161.
- Kleiman, E. M., Turner, B. J., Fedor, S., Beale, E. E., Huffman, J. C., & Nock, M. K. (2017). Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of abnormal psychology, 126*(6), 726.
- Klimstra, T. A., & Denissen, J. J. (2017). A theoretical framework for the associations between identity and psychopathology. *Developmental psychology, 53*(11), 2052.
- Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods, 53*(6), 2576-2590.
- Kurek, A., Jose, P. E., & Stuart, J. (2017). Discovering unique profiles of adolescent information and communication technology (ICT) use: Are ICT use preferences associated with identity and behaviour development?. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 11*(4).
- Lavery, M. R., Acharya, P., Sivo, S. A., & Xu, L. (2019). Number of predictors and multicollinearity: What are their effects on error and bias in regression?. *Communications in Statistics-Simulation and Computation, 48*(1), 27-38.
- Lenhart, A., Duggan, M., Perrin, A., Stepler, R., Rainie, H., & Parker, K. (2015). Teens, social media & technology overview 2015. *Pew Research Center* [Internet & American Life Project].
- Loveys, K., Torrez, J., Fine, A., Moriarty, G., & Coppersmith, G. (2018). Cross-cultural differences in language markers of depression online. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 78-87).

- Luyckx, K., Klimstra, T. A., Duriez, B., Van Petegem, S., & Beyers, W. (2013). Personal identity processes from adolescence through the late 20s: Age trends, functionality, and depressive symptoms. *Social Development, 22*(4), 701-721.
- 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.
- Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of personality and social psychology, 3*(5), 551.
- Marden, J. I. (2004). Positions and QQ plots. *Statistical Science, 606-614*.
- Meeus, W. (2011). The study of adolescent identity formation 2000–2010: A review of longitudinal research. *Journal of research on adolescence, 21*(1), 75-94.
- Meeus, W., van de Schoot, R., Keijsers, L., & Branje, S. (2012). Identity statuses as developmental trajectories: A five-wave longitudinal study in early-to-middle and middle-to-late adolescents. *Journal of Youth and Adolescence, 41*(8), 1008-1021.
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology, 84*(4), 857.
- Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAI) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine, 52*(6), 446-462.
- Nanomi Arachchige, I. A., Sandanapitchai, P., & Weerasinghe, R. (2021). Investigating machine learning & natural language processing techniques applied for predicting depression disorder from online support forums: A systematic literature review. *Information, 12*(11), 444.
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse processes, 45*(3), 211-236.
- Nobles, A. L., Glenn, J. J., Kowsari, K., Teachman, B. A., & Barnes, L. E. (2018). Identification of imminent suicide risk among young adults using text messages. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-11).
- Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-Injurious Thoughts and Behaviors Interview: development, reliability, and validity in an adolescent sample.

- O'dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. *Internet Interventions*, 2(2), 183-188.
- Ollendick, T. H., King, N. J., & Muris, P. (2002). Fears and phobias in children: Phenomenology, epidemiology, and etiology. *Child and Adolescent Mental Health*, 7(3), 98-106.
- Ormel, J., Oerlemans, A. M., Raven, D., Laceulle, O. M., Hartman, C. A., Veenstra, R., ... & Oldehinkel, A. J. (2017). Functional outcomes of child and adolescent mental disorders. Current disorder most important but psychiatric history matters as well. *Psychological Medicine*, 47(7), 1271-1282.
- Osborne, J. (2002). Notes on the use of data transformations. *Practical assessment, research, and evaluation*, 8(1), 6.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296.
- 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.
- Pérez-Torres, V., Pastor-Ruiz, Y., & Abarrou-Ben-Boubaker, S. (2018). Youtuber videos and the construction of adolescent identity. *Comunicar. Media education research journal*, 26(1).
- Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry*, 56(3), 345-365.
- Portes, P. R., Sandhu, D. S., & Longwell-Grice, R. (2002). Understanding adolescent suicide: A psychosocial interpretation of developmental and contextual factors. *ADOLESCENCE-SAN DIEGO-*, 37, 805-814.
- Ramgoon, S., Bachoo, S., Patel, C., & Paruk, Z. (2006). Could a healthy ego identity serve as a protective factor against suicidal tendencies? A pilot study. *Journal of Child and Adolescent Mental Health*, 18(2), 49-54.
- Rich, G. J. (2003). The positive psychology of youth and adolescence. *Journal of Youth and Adolescence*, 32(1), 1.
- Ruch, D. A., Sheftall, A. H., Schlagbaum, P., Rausch, J., Campo, J. V., & Bridge, J. A. (2019). Trends in suicide among youth aged 10 to 19 years in the United States, 1975 to 2016. *JAMA network open*, 2(5), e193886-e193886.

- Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion, 18*(8), 1121-1133.
- Schwartz, S. J., Klimstra, T. A., Luyckx, K., Hale III, W. W., Frijns, T., Oosterwegel, A., ... & Meeus, W. H. (2011). Daily dynamics of personal identity and self-concept clarity. *European Journal of Personality, 25*(5), 373-385.
- Seabrook, E. M., Kern, M. L., Fulcher, B. D., & Rickard, N. S. (2018). Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. *Journal of medical Internet research, 20*(5), e168.
- Shen, J. H., & Rudzicz, F. (2017, August). Detecting anxiety through reddit. *In Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality* (pp. 58-65).
- Sonnenschein, A. R., Hofmann, S. G., Ziegelmayer, T., & Lutz, W. (2018). Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. *Cognitive behaviour therapy, 47*(4), 315-327.
- Stamatis, C. A., Meyerhoff, J., Liu, T., Sherman, G., Wang, H., Liu, T., ... & Mohr, D. C. (2022). Prospective associations of text-message-based sentiment with symptoms of depression, generalized anxiety, and social anxiety. *Depression and anxiety, 39*(12), 794-804.
- Thelwall, M., Thelwall, S., & Fairclough, R. (2021). Male, Female, and Nonbinary Differences in UK Twitter Self-descriptions: A Fine-grained Systematic Exploration. *Journal of Data and Information Science, 6*(2), 1-27.
- Throuvala, M. A., Griffiths, M. D., Rennoldson, M., & Kuss, D. J. (2019). Motivational processes and dysfunctional mechanisms of social media use among adolescents: A qualitative focus group study. *Computers in Human Behavior, 93*, 164-175.
- Tølbøll, K. B. (2019). Linguistic features in depression: a meta-analysis. *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift, 4*(2), 39-59.
- van Doeselaar, L., Klimstra, T. A., Denissen, J. J., Branje, S., & Meeus, W. (2018). The role of identity commitments in depressive symptoms and stressful life events in adolescence and young adulthood. *Developmental psychology, 54*(5), 950.
- Vijendra, S., & Shivani, P. (2014). Robust outlier detection technique in data mining: A univariate approach. *arXiv preprint arXiv:1406.5074*.
- Watkins, E., & Teasdale, J. D. (2004). Adaptive and maladaptive self-focus in depression. *Journal of affective disorders, 82*(1), 1-8.

Weems, C. F., Silverman, W. K., & La Greca, A. M. (2000). What do youth referred for anxiety problems worry about? Worry and its relation to anxiety and anxiety disorders in children and adolescents. *Journal of abnormal child psychology*, 28, 63-72.

Young, J., Bishop, S., Humphrey, C., & Pavlacic, J. M. (2023). A review of natural language processing in the identification of suicidal behavior. *Journal of Affective Disorders Reports*, 100507.