Theses and Dissertations   Graduate School

2024

USING NATURAL LANGUAGE PROCESSING TO UNDERSTAND THE LIVED EXPERIENCES OF PEOPLE IDENTIFYING WITH ADHD: WHAT THEMES EMERGE IN SOCIAL MEDIA POSTS?

Gabby C. Scalzo
Virginia Commonwealth University

USING NATURAL LANGUAGE PROCESSING TO UNDERSTAND THE LIVED
EXPERIENCES OF PEOPLE IDENTIFYING WITH ADHD: WHAT THEMES EMERGE IN
SOCIAL MEDIA POSTS?

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctoral of
Philosophy in Clinical Psychology at Virginia Commonwealth University

By: GABRIELLA C. SCALZO

B.S., Virginia Polytechnic Institute and State University, 2017

M.S., Virginia Commonwealth University, 2019

Director: Jared Keeley, Ph.D.

Associate Professor of Psychology

Department of Psychology

Virginia Commonwealth University

Richmond, Virginia

April 2024

Acknowledgements

I am eternally grateful to the many people who helped me make this dissertation possible. Firstly, thank you to my advisor, Dr. Jared Keeley, for his mentorship, compassion, and steadfastness. You were my rock. Thank you to my first advisor, Dr. Scott Vrana, whose belief in and support of me has continued far beyond his well-deserved retirement, and who served on this committee. I also thank my former committee member, Dr. Joshua Langberg, and my current committee members, Drs. Bridget Thomson-McInnes, Dace Svikis, and Bryce McLeod, for their valuable contributions and support. Thank you to my research assistants, Mallory and Waverely, whose support and hard work were irreplaceable. This dissertation is dedicated to Graziella and Dominick Scalzo, the first people to ever believe in me; my life partner, Amory; and my friends, Wendy, Elliott, Sam, Elise, and Jaime. Thank you forever. Finally, immense thanks to my cat, Henrietta, whose contributions to this project were vocal and numerous.

Table of Contents

List of Figures

USING NATURAL LANGUAGE PROCESSING TO UNDERSTAND THE LIVED
EXPERIENCES OF PEOPLE IDENTIFYING WITH ADHD: WHAT THEMES EMERGE IN
SOCIAL MEDIA POSTS?

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctoral of
Philosophy in Clinical Psychology at Virginia Commonwealth University

By: GABRIELLA C. SCALZO

B.S., Virginia Polytechnic Institute and State University, 2017

M.S., Virginia Commonwealth University, 2019

Director: Jared Keeley, Ph.D.

Professor of Psychology

Department of Psychology

Abstract

Compared to the amount of research conducted on how to identify and understand
children with Attention-Deficit Hyperactivity Disorder (ADHD), there has been relatively little
work done to understand the lived experiences of adults with ADHD. Increased understanding of
how adults with ADHD conceptualize themselves in the context of their diagnosis would help
clinical experts tailor research and treatments to better serve these communities. However, there
are several barriers towards conducting high-quality qualitative research, including time- and
labor-intensity. This study, informed by qualitative research traditions, used innovative data
sources (i.e., social media) and analytic techniques (i.e., machine learning) to reduce these
barriers to research and open a window to understand how adults identifying with ADHD talk
about their lived experiences. The results of the analysis were compared to the conceptualization
of ADHD found in the Diagnostic and Statistical Manual, 5th Edition (DSM-5) and in a seminal

handbook geared towards clinicians. Twenty-three unique themes were identified, which

addressed a variety of topics including primarily inattentive-related symptoms, interpersonal-

related concerns, topics related to medication, and topics related to broader systems and formal

supports. There was moderate agreement between the identified themes and the *DSM-5* and high

agreement between the identified themes and the gold standard treatment manual.

Introduction

While extensive research has been conducted on how to best identify children with ADHD, as well as to interview caregivers of children with ADHD (e.g., Peterson et al, 2024; Harborne, Wolpert, & Clare, 2004), comparatively less work has been done to study the experiences and perspectives of people with ADHD in adulthood. Listening to the perspectives of people with ADHD can provide us with valuable insights into how the disorder affects them over the lifespan—especially in the cases of people with other marginalized identities, who have been underrepresented in research. Relatedly, it can help us understand potential differences between how people with ADHD conceptualize their own experiences and how clinical experts conceptualize these experiences. This understanding can have valuable implications for future research and clinical practice by helping experts identify gaps in their knowledge, develop new or adjusted treatments, and better connect with the populations they seek to serve.

There are several barriers to conducting such research. It can be difficult to recruit participants, especially in high volumes, and traditional qualitative methods for analyzing the lived experiences of a group, such as phenomenological studies, are time- and labor-intensive (Sloan & Marx, 2004). As with all research strategies, researchers must be careful not to introduce bias into their analyses; this is particularly salient when interviewers' identities influence how they interact with participants, how participants choose to engage with interviewers, and how coders interpret qualitative data to codify themes (Law et al., 1998). Furthermore, if experts in the field are seeking to understand how their language and conceptualizations differs from participants' own, it may be particularly difficult to avoid imposing a pre-existing lens on the data analysis.

Several of these barriers can be addressed by utilizing advancements in the field of computer science. Firstly, the use of "big data" such as publicly available social media posts can reduce the need for participant recruitment and provide access to an otherwise unobtainable amount of data; computer science tools can turn data collection and cleaning of this kind into a relatively automated process, saving significant labor hours. Second, machine-learning tools which engage in natural language processing (NLP) have been developed to analyze the semantic content of human language (e.g., TopEx; Olex et al., 2020; Olex et al., 2021). These programs can be trained on large volumes of text to "learn" human language in a bottom-up fashion, which is similar to how humans learn language via immersion. In both cases, the meaning of words is derived contextually by understanding how words are used in relation to each other, rather than top-down approaches to learning or analyzing language, which involve an external source assigning meaning. Because of this, using NLP can address issues of labor- and time-intensity, as well as potentially reduce bias in data analysis if the NLP is appropriately trained.

While the use of NLP is becoming more popular across disciplines and technology continues to improve, the use of NLP in clinical psychology is still relatively unexplored. This study is an initial exploration of using NLP to compare the self-conceptualizations of individuals identifying with ADHD to the conceptualizations of clinical experts, with the goal of better understanding how to bridge the gap between clinical experts and the populations they serve.

Literature Review

Attention-Deficit Hyperactivity Disorder

Attention-Deficit Hyperactivity Disorder (ADHD) is categorized in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) as a neurodevelopmental disorder, meaning that it develops early in childhood, results in developmental concerns in multiple domains such as academic and social functioning, and is likely to maintain throughout the lifespan (American Psychological Association, 2013). In particular, ADHD is characterized by the DSM-5 as involving "impairing levels of inattention, disorganization, and/or hyperactivity-impulsivity" (American Psychological Association, 2013). ADHD may present with Predominantly Inattentive symptoms, Predominantly Hyperactive symptoms, or Combined; the presentation type may change as the individual ages and/or their life circumstances change (American Psychological Association, 2013). Notably, ADHD is also associated with a variety of other features. Many people with ADHD struggle with difficulties in emotion regulation, broader executive functioning, and social functioning (American Psychological Association, 2013). People with ADHD are also at considerably elevated risk for suicidal ideation, risk-taking behaviors, and suicide attempts (American Psychological Association, 2013).

Historically, a significant portion of ADHD research was conducted on children (or the caregivers of children) in White North American and Western European cultures; in fact, the DSM criteria for diagnosing ADHD were originally centered entirely around childhood-specific behaviors and later amended to include manifestations in adulthood (Williamson & Johnston, 2015). However, many adults are still affected by ADHD and their presentations can vary significantly from childhood presentations (Williamson & Johnston, 2015). Concerns in adulthood related to ADHD may include social rejection, difficulties in obtaining or maintain desired employment, and elevated levels of interpersonal conflict (Kessler et al., 2006; Harpin et al., 2016). Impulsivity also remains a concern for many adults, and can manifest in a variety of

ways, including issues with substance use and risk-taking (e.g., unsafe sex, reckless driving; Turgay, et al., 2012). ADHD is also associated with the development of other mental disorders in adulthood, such as certain personality disorders and trauma-related disorders (Ottosen et al., 2019).

Social difficulties and interpersonal conflict are two associated features of ADHD not included in the diagnostic criteria (though they are cited elsewhere in the DSM-5; American Psychological Association, 2013). Many individuals with ADHD endorse difficulty making friends and suffering bullying from their peers (Canu & Carlson, 2007). Research suggests they may also be particularly sensitive to rejection, which exacerbates the effects of these concerns (Babinski et. al., 2019; Bondü & Esser, 2015). Because a large body of research shows that lack of strong social support is affiliated with a wide variety of negative outcomes (for an overview, see Turner & Brown, 2010), it is particularly important to understand the extent to which a person with ADHD experiences these associated features.

The Need for Qualitative Research

People identifying with ADHD, especially those who were not diagnosed in childhood and/or did not have consistent access to formalized supports, may have very different conceptualizations of their symptoms than experts in the field, and in particular their most salient concerns may not align with the diagnostic criteria outlined by diagnostic manuals such as the DSM-5 or even treatment manuals which are geared towards clinical professionals with the intention of addressing ADHD symptoms in clients. In cases such as this, it is particularly important to research the lived experiences of the population in question. Qualitative research and, more specifically, phenomenological research, is an invaluable tool for better understanding the perspectives of the people the mental health field seeks to help.

"Qualitative research" is a broad term which can describe a wide variety of research methodologies. Broadly, it refers to the practice of analyzing non-numerical data, with examples including in-depth interviews, certain kinds of observational data, and historical research (i.e., analyzing past events; Fossey et al., 2002). Qualitative research is particularly helpful for increasing understanding of complex phenomena such as animal or human behavior, internal experiences, and areas of new understanding (Fossey et al., 2002). It is also an excellent method for increasing community engagement in research. Because qualitative research often involves asking participants to share their own experiences, perspectives, and beliefs, it is well-suited for prioritizing client-centered care as well.

Phenomenology is a specific type of qualitative research which is particularly well-suited for understanding the lived experiences of a group. Per Law et al. (1998, pg. 2), phenomenology "answers the question: 'what is it like to have a certain experience?'" Meanings of "experience" may include broader concepts such as an identity group, "base" experiences such as a particular emotion, or specific experiences such as "being a first-year graduate student in clinical psychology." Researchers taking a phenomenological approach assume that there is an "essence" to these kinds of shared experience and seek to understand common underpinnings (Law et al., 1998). The self is used as a tool to understand the shared experiences of others (Law et al., 1998). Methods of data collection may include interviews, written accounts, and even artistic mediums such as poetry or illustration (Law et al., 1998).

Pre-existing Qualitative Studies on ADHD Adults

A review of the literature shows that previous qualitative research studies which focused on the lived experiences of people with ADHD have identified a variety of topics which are salient to these individuals regarding their diagnosis. Many participants across studies did

identify themes related to prototypical ADHD core symptoms such as inattentiveness, hyperactivity, and impulsivity (Gully, 2009; Henry & Jones, 2011; Michielsen et. al., 2018; Osagiede-Omon, 2020; Schaedler, 1999; Waite & Tran, 2010; Zwennes & Loth, 2019). Researchers across studies have also found that issues of interpersonal conflict, social isolation, and feeling "different" from the people around them were common concerns for people with ADHD (Henry & Jones, 2011; Michielsen et. al., 2018; Osagiede-Omon, 2020; Stenner, O'Dell, & Davies, 2018; Waite & Tran, 2010). Relatedly, multiple participants across these studies described themselves as "rebellious" and more likely to speak out against perceived injustice, and women in particular acknowledged that they were more outspoken than was typical for their gender in their generation (Henry & Jones, 2011; Michielsen et. al., 2018).

Feelings of being "othered" and having social difficulties tended to decrease for participants over time, especially after they left the education system and grew into adulthood; however, some participants across multiple studies reported that their social difficulties endured (Henry & Jones, 2011; Michielsen et. al., 2018). Themes related to low/compromised self-esteem were also common. Many participants across studies said that they had struggled to understand their behavior and typically experienced feelings of shame and guilt for "underachieving" (Gully, 2009; Michielsen et. al., 2018; Schaedler, 1999; Stenner, O'Dell, & Davies, 2018; Waite & Tran, 2010; Zwennes & Loth, 2019). These feelings were typically reinforced by authority figures in these participants' lives (e.g., Stenner, O'Dell, & Davies, 2018; Zwennes & Loth, 2019).

In several studies, participants also described the benefits of receiving an ADHD diagnosis, which typically focused on how understanding their ADHD helped them view themselves in a more positive, forgiving light (Henry & Jones, 2011; Stenner, O'Dell, & Davies,

2018; Waite & Tran, 2010; Zwennes & Loth, 2019). In other words, access to a diagnosis was a protective factor against feelings of shame and low self-worth (e.g., Osagiede-Omon, 2020; Stenner, O'Dell, & Davies, 2018). This result is particularly imperative because another common theme among participants is that they were often not diagnosed with ADHD until adulthood (sometimes late adulthood, e.g., Henry & Jones, 2011) and some had still not been formally diagnosed (e.g., Michielsen et. al., 2018). Participants across multiple studies expressed the belief that their lives would have been improved if they had been diagnosed with ADHD as children (Stenner, O'Dell, & Davies, 2018; Zwennes & Loth, 2019).

While some groups of participants described helpful experiences in therapy (Henry & Jones, 2011; Stenner, O'Dell, & Davies, 2018), the participants across several studies emphasized their desire for better services, either having received little to no formal support or having found the treatments available lacking (Osagiede-Omon, 2020; Waite & Tran, 2010; Zwennes & Loth, 2019). Notably, both studies which focused on people of color as the intersectional identity of interest highlighted themes involving lack of access to formal supports (Osagiede-Omon, 2020; Waite & Tran, 2010) which is in-keeping with research that suggests Western society particularly disadvantages people of color in terms of access to mental health services (Morris & Perry, 2017; Osagiede-Omon, 2020).

In contrast to the disparities in access to formal supports, participants across the majority of studies tended to report on coping skills they had developed outside of formal therapeutic contexts. While the coping mechanisms themselves varied (e.g., sometimes participants emphasized their decision to seek external structure while other groups of participants sought out environments where structure was not required), participants almost exclusively reported having learned these coping mechanisms "for themselves" (Gully, 2009; Henry & Jones, 2011;

Schaedler, 1999; Waite & Tran, 2010). Some groups of participants also discussed the influence of parents or other mentors in teaching them helpful coping strategies (Osagiede-Omon, 2020). Participants across several studies also endorsed histories of advocating for themselves or others (e.g., "sticking up for the underdog" in Henry & Jones, 2011; parents advocating for their children in Stenner, O'Dell, Davies, 2018) and working to tailor their environments to be beneficial to them (Gully, 2009; Henry & Jones, 2011; Schaedler, 1999).

Barriers to Conducting Qualitative Research

While qualitative research has many benefits, there are several important barriers against conducting this type of work. Traditionally, qualitative research has involved time- and labor-intensive work on several fronts. First, it is well-lamented in the field that recruiting research participants is a difficult process, whether hundreds of participants are required for a quantitative study or a handful of participants are required for in-depth interviewing. Second, much phenomenological research requires trained interviewers to spend hours per participant conducting semi-standardized interviews. After the interviews have been collected, best practice is for multiple coders to carefully review the interviews to identify recurring themes within and between narratives. At every step in the process, researchers must be aware of how their own biases, including their personal identities, influence their process (Law et al., 1998).

Because coding data is done manually and is inherently subjective, based on the specific researchers who are interpreting the data, it is particularly difficult to compare the results of different but related studies, impairing the ability to build a larger knowledge base. Furthermore, because studies of this kind take substantial time and resources to complete, progress in advancing the field can be slow.

This dissertation is informed by phenomenological research principles and uses both innovative methodology (i.e., natural language processing) and data sources (i.e., social media posts) to address current barriers to research and further the field's understanding of the lived experiences of people with ADHD. The remainder of the literature review provides more detail on the advantages of my techniques.

ADHD Self-Advocacy and the Importance of Social Media

There is a growing self-advocacy movement for people identifying with ADHD to offer intracommunity support and raise awareness of their lived experiences. A significant portion of this advocacy takes place in online spaces on social media websites, and there is much to be learned from self-advocates, who are already actively seeking to communicate answers to many of the questions we are asking as researchers interested in the community's experiences.

Publicly available posts from ADHD advocates feature efforts to raise awareness and, perhaps more importantly, acceptance on a variety of topics related to ADHD. These can range from posts validating ADHD experiences (e.g., hyperfixation; so1987, 2021) to highlighting social justice issues such as gender differences in diagnosis (adhdbreadbin, 2020) or calling attention to racism within the neurodivergent community (ablednt, 2020). Some advocates also express frustration with how ADHD is described and diagnosed (e.g., alienjuuce, 2021). A popular hashtag for discussing ADHD issues on social media platforms such as Tumblr and Twitter is "#actuallyADHD."

In addition to raising awareness, there are online spaces dedicated to disseminating coping skills; self-advocates tend to share strategies they have developed for themselves and/or learned from others in the hopes that fellow community members may benefit from them (e.g.,

"tips for dealing with sensory issues…" by biggest-gaudiest-patronuses, 2021; tips for dental hygiene, Badger, 2021). This advice can be disseminated proactively, but online communities are also a forum for individuals to request advice on specific topics (e.g., Potential-Search-376, 2022). Given the significant barriers to receiving an ADHD diagnosis and subsequent treatment that many people face (Dumit, 2006; Stenner et. al., 2019) these sources of informal psychoeducation are valuable. Many popular content creators who focus on ADHD issues structure their content in ways which are considered more accessible for people with ADHD and generally more easily consumed by the public. For example, the creator of a popular web-comic titled "ADHD Alien" uses a cartoonish illustration style and metaphorical storytelling to illustrate their struggles and successes with ADHD (for examples of the creator's work, see ADHD-alien, 2020a, 2020b).

While much of the advice shared in online communities does not formally reference evidence-based strategies for coping with ADHD, many of the coping strategies being shared by advocates do mirror the evidence base (e.g., breaking large tasks into less complex ones, biggest-gaudiest-patronuses, 2021; "coping thoughts," thevirgodoll, 2021). In addition, users may share "life hack"-style advice, such as keeping multiple laundry hampers in different locations around the living space to address executive functioning issues which may otherwise lead to clothes being scattered on the floor. Notably, much of this advice encourages those with ADHD to relax or forgo societal norms around what a "clean" or "successful" living space looks like. Instead, readers are encouraged to do "what works for them." These coping strategies are in-line with ones which a client-centered therapist operating from an evidence-based approach would recommend.

Of course, it should be noted that the commentaries referenced above are a small sample of the population and may not necessarily be representative, neither of ADHD self-advocates nor people identifying with ADHD more broadly. Posts curated under certain social media communities, such as specific hashtags, encourage a selection bias. Nonetheless, it is important to consider the perspectives there within—especially given that such populations are currently underrepresented in formal research. This study evaluates posts from an online community for people identifying with ADHD which is hosted on the popular social media website, Reddit, as one avenue for gaining wisdom from the ADHD community.

There are several advantages to using Reddit posts as a data source. First, a large volume of data can be collected this way. Second, analyzing social media posts allows access to data that is not influenced by the presence of an interviewer; people are likely to word posts addressed to their peers differently than they would word responses to someone they perceive as holding a position of power or elevated social status (i.e., a scientist). Third, the posts are a mix of historical reflection (e.g., a post author reflecting on being bullied as a child) and immediate concerns (e.g., a post author asking for ADHD-friendly tips to get in the habit of teeth-brushing). This mixture provides access to a wide range of thematic topics.

Natural Language Processing as a Data Analysis Strategy

As mentioned previously, natural language processing (NLP) techniques are a subset of machine-learning, related to the concept of artificial intelligence (AI). A machine-learning tool is one which uses computer science techniques to "train" a software program by exposing it to data from which it "learns." NLPs in particular are used to analyze human language. In fact, one early NLP called latent semantic analysis (LSA) was originally developed as a theory of how humans learn language (Landauer, Foltz, & Laham, 1998; Landauer, 2007). Generally speaking, NLPs

use a large volume of text to create associations between words in order to "learn" what the meaning of each word is in relation to the other words. This is important because it emphasizes that a typical NLP learns in a bottom-up fashion; rather than humans deciding how certain words should be defined or relate to each other, a user simply provides exposure to human language and allows the NLP to determine how words are interconnected to form meaning (Landauer, 2007). This is similar to how humans can learn a new language by "immersion;" over time, a person learns which words are used together in which contexts, and therefore figures out the meanings of each word.

A variety of NLPs have been developed by computer scientists, most of which require significant expertise in computer science to use successfully, which has been a barrier to implementing NLP in other disciplines. The specific NLP technique used in this study is a publicly available application called TopEx (Olex et al., 2020; Olex et al., 2021). TopEx was specifically developed to be more accessible to users from non-computer science disciplines by eliminating the need for prior experience with NLPs or programming more broadly; however, it is still valuable to understand in greater detail how TopEx works.

TopEx is domain agnostic, meaning that it could be applied to a data source of the user's choice without modifying the program (Olex et al., 2021). TopEx can also be "seeded" with suggested topics or semantic connections, based on user interest. This may be particularly relevant if users want to highlight or prioritize certain known topics in a dataset, or to help guide manual analysis of the results. For example, Olex and colleagues (2020) had medical experts seed an analysis of the reflective writing of medical students based on known challenges which students are understood to face, in order to help identify the extent to which these challenges were present in the students' writing.

The ability to seed TopEx introduces a trade-off: it adds a potential source of bias, as this is a "top-down" component which involves researchers preemptively identifying themes, but is helpful in guiding the data analysis. It is also important to note that, because of how NLPs are constructed, seeding an analysis does not prevent new themes from being discovered as the software will continue to search for other relevant connections within the data. This is different than when researchers identify themes with which to code during manual qualitative analysis, because human coders may have the subconscious tendency (or even be consciously encouraged) to "find only what they are looking for," (i.e., fall prey to confirmation bias).

Once given a corpus of text, TopEx will create a matrix that represents how often words occur in relation to each other and from where they occurred. For example, the word *bacteria* may frequently appear in a biology textbook, but significantly less frequently appear in a physics textbook, and these disparities will be accounted for in the matrix. *Bacteria* would also be likely to co-occur in a passage that uses the term *organism*; in contrast, *organism* will be unlikely to appear in the physics textbook, and these patterns of occurrence also contribute to how TopEx learns about contextual meaning.

The analysis does not stop with a direct measure of co-occurrence. This is because words that are highly synonymous may have high levels of semantic similarity but not occur together in a single text (Landauer, 2007). For example, *doctor* and *physician* are highly related words that would likely not occur in close proximity, because they serve nearly identical functions in a passage. To expand on the previous example, the analysis would consider that *doctor* and *physician* themselves may not frequently co-occur, but would both co-occur with words such as *stethoscope* and *examination,* and therefore learn that *doctor* and *physician* are semantically

similar. It is within this web of contextual relationships that TopEx evaluates the meaning of a passage based on the meaning of the individual words.

Like other NLPs (e.g., Foltz, 2007), TopEx's evaluation of meaning may function similarly to humans' by focusing on the concept of *gist.* That is, both NLPs and human language users focus on the overall conceptualized meaning of a passage by analyzing what, in general, a passage communicated (Foltz, 2007). TopEx specifically identifies the main theme or topic of a sentence by identifying "most informative phrases," which are based on which words are determined to have greater mathematical weight in the matrix. This method assumes that the meaning of a sentence, and by extension a longer passage, is a function of the meanings of the individual words. To put it simply, the whole is the sum of its parts. TopEx can create word maps based on its identified themes by highlighting the most important and frequently used words both within and across individual sources. It can then create topic clusters in a two-dimensional space, helping researchers visualize their results. For example, Olex and colleagues (2021) used TopEx to identify trends in publicly available Twitter "tweets" related to COVID-19. Clustered topics included "staying home safe," "COVID Testing," and "Media responds to Trump" (Olex et al., 2021).

The present study used TopEx to analyze the themes present within social media posts on a Reddit forum which is dedicated to people identifying with ADHD. TopEx was used to visualize major themes in a less biased and labor-intensive way compared to traditional qualitative methods, while retaining the spirit of qualitative research by focusing on the lived experiences of people identifying with ADHD in their own words. These themes are then compared to how ADHD is described by clinical experts in the DSM-5 and in a seminal handbook geared towards clinicians, described in more detail shortly.

Statement of the Problem

The field of clinical psychology would benefit from continued research into how individuals identifying with ADHD conceptualize their own presenting concerns. This research would provide greater insights into the lived experiences of the people our work seeks to support and help both researchers and clinicians better connect with their population of interest. However, historical barriers to conducting this research have included the difficulty of recruiting participants, the time- and labor-intensity of data analysis and collection, and the risk of introducing bias into the analysis.

The present study addresses several of these barriers using innovative data collection and analysis strategies. A large volume of data was collected by gathering social media posts from Reddit, a popular social media platform which has a forum community ("r/ADHD") dedicated to people identifying with ADHD. People identifying with ADHD use this "subreddit" to talk about their concerns and offer both advice and support to others within their community. Then, an NLP program called TopEx was used to analyze major themes present in the Reddit posts to gain an understanding of what issues are most salient to people identifying with ADHD. Finally, these themes are compared to both the DSM-5 criteria for diagnosing ADHD and a widely-regarded handbook on diagnosing and treating ADHD to determine the extent to which these conceptualizations align with experts' descriptions of the disorder. The objectives of the present study were as follows: (a) to understand how people identifying with ADHD on a popular social media site conceptualize their most salient concerns related to their diagnosis, (b) to investigate the extent to which people identifying with ADHD's own language and conceptualization regarding ADHD aligns with how experts conceptualize ADHD, and (c) to provide preliminary

support for the use of NLP, specifically TopEx, as a broader methodology in the field of clinical psychology.

Hypotheses

Given the large volume of data and the assumed diversity of the dataset, it was expected that a wide variety of topics would be identified by TopEx. In particular, based on the recurring themes identified by previous qualitative studies and an anecdotal review of self-advocacy community discussions on similar social media platforms, I hypothesized that the following topics would feature prominently in the results: (a) experiences of social rejection, (b) medication concerns, including struggling with gaining access to medication, (c) barriers to obtaining a clinical diagnosis, (d) seeking advice on coping strategies, and (e) celebrations of successes and/or uplifting positive aspects of having ADHD.

Method

Data Source

Data was collected by "scraping" social media posts from a specific forum (called a "subreddit") on the popular social media website, Reddit. Scraping is a computer science technique which imports information from a website into a local file, such as a spreadsheet or .txt files. In this case, the texts of posts on r/ADHD, a subreddit that describes itself as "an inclusive, disability-oriented peer support group for people with ADHD," was scraped for analysis (r/AHDD, 2008). r/ADHD is a moderated community of approximately 1.5 million people (at the time of data collection; r/ADHD, 2008) which hosts posts in a variety of self-labelled categories, such as "Tips/Suggestions" and "Seeking Empathy/Support." Posts were collected regardless of category in order to fully capture the range of experiences being discussed by members of the

community. The data from r/ADHD was scraped on June 27, 2022 and all unique, non-deleted posts made in the subreddit's history were scraped.

It is important to define what was counted as a "post" for analysis. Each post on r/ADHD is created as a separate discussion thread with an author, title, and textual body expanding on the topic introduced by the title. Each post is associated with a unique hyperlink, which is how the scraping process separates the forum's contents into unique posts. Similarly, each comment in an individual discussion thread is associated with an author and a distinct hyperlink, so comments may be separately identified as well. For the purposes of this analysis, however, the analysis was limited to the "original" post itself, meaning that each data point consisted of a thread title plus the textual body of the post. Follow-up comments by other users were excluded. These comments were excluded for parsimony and to reduce potential noise in the data which would otherwise be exacerbated by "off-topic" tangents in the replies to the original post. The scraping process returned 236,123 unique posts.

Another potential concern in standardizing the data was how to approach multiple posts created by the same user. While 236,123 unique posts were returned by the scraping process, this is not equivalent to having 236,123 unique *participants*. Due to the reported membership of the community totaling approximately 1.5 million users, it can be reasonably assumed that enough unique participants contributed to the overall posting volume to create a sufficient sample size for analysis, especially in contrast to the number of participants who are traditionally interviewed for qualitative work (with individual studies typically interviewing less than twenty people and sometimes as few as two to three). However, it is possible that frequent posters—people who may post multiple times a week, for example, in interest of social interaction, higher-intensity help-seeking, etc.—may slightly skew the data by inflating the frequency of certain themes.

The decision was made *not* to filter for repeat posters in the interest of collecting a richer dataset, and because the normative practice in a subreddit such as r/ADHD is to make individual posts for each relevant concern, even if the user has multiple salient issues to discuss over a given period of time. For example, if a user wanted to discuss multiple issues which occurred in their past but were distinct from each other (e.g., topics regarding childhood bullying and difficulties dating as an adolescent), they would likely make separate posts for each topic. Users are also more likely to make a post only when a given topic is relevant. For example, a user who has a history of struggling in their sex life might only choose to post about this topic after a recent unfulfilling sexual encounter. This would be in contrast to a traditional qualitative dataset, where each participant undergoes an extensive interview all at once and is asked to speak to all of their concerns at one time regardless of volume or immediacy. Therefore, by including all of the posts made by any user, the dataset can be better likened to the nature of an extensive qualitative interview, where multiple salient concerns may arise in any order.

Procedure

After the raw data was collected, it was analyzed using TopEx. Because one goal of the present study was to compare how people identifying with ADHD talk about their experiences to how experts talk about ADHD, TopEx was not seeded with any predictive themes. When entering the data into TopEx, it was discovered that the website application which hosted the TopEx programming had a processing capacity of approximately 500 posts at a time. Thus, random subsamples were drawn from the total dataset, composed of 500 posts each. Each subset was analyzed using TopEx and coded by two coders independently. Subsamples were created and analyzed until saturation was accomplished after 5 subsamples. Due to saturation being

reached, it can be reasonably assumed that these subsamples are representative of the whole sample.

While demographic information about specific participants or the registered members of r/ADHD was not unable, some information about the general demographics of Reddit's broader userbase has been publicly reported. It is important to note that these overall statistics may not be representative of the subset of Reddit users who frequent r/ADHD specifically. A survey conducted in 2017 indicated that approximately 67% of Reddit users identified as male. 64% of the userbase that responded to the survey was between the ages of 18 and 29; 29% reported being between the ages of 30 and 49, 6% reported being between 50 and 64 years old, and 1% of users reported being over the age of 65. Reliable and inclusive data regarding racial/ethnic identities of users was not readily available (Sattleburg, 2021).

Data Analysis

TopEx produces two methods of visualizing data, both of which were used in the analysis: a two-dimensional scatter plot which visualized "clusters" of sentences which contained similar most informative phrases, and one word cloud per cluster which visualized the frequencies of the individual words which comprise the cluster's most informative phrases. These strategies in combination facilitated the identification of themes. TopEx creates interactive visualizations, wherein each point on the scatterplot can be hovered over to produce the full text of the sentence being represented, along with the most informative phrase. This, paired with the word clouds created for each cluster, elucidated the most frequently used words within the cluster's most informative phrases. The number of clusters generated for each subsample was able to be manually adjusted to increase goodness of fit of cluster number. The number of clusters per subsample was selected by the same coder for each subsample based on exploratory

analysis by examining the cluster breakdown produced at different prescribed cluster numbers (i.e., the coder tested cluster generation at different numbers of prescribed clusters and did an initial evaluation if clusters were either too numerous and therefore redundant or too few and therefore noisy/imprecise). For four of the five subsamples, the optimal number of clusters was determined to be 15. For the remaining subsample, the optimal number was 10.

Clusters were assigned theme names based on the content of the most informative phrases and correlated word cloud within each cluster by constructing the simplest intelligible sentence fragment which can be produced from a cluster's word cloud. However, in some cases further context was required to ensure the sentence fragment was being constructed accurately. For example, a cluster whose word cloud contains the words "bullied" and "kid" could produce either "bullied as a kid" or "bullied another kid" which would result in significantly different interpretations. The ability to easily visualize the full sentences from which most important phrases were derived was thus used to verify the context of the word cloud. A second coder was used to help reduce potential bias in this process; the first and second coders independently labeled each cluster and compared results.

Across five subsamples, initial agreement between the two coders within a subsample ranged from 80% to 100%, with average agreement being 89.4%. After identifying which cluster labels were not in agreement, the two coders and myself held a meeting to reconcile differences in cluster labels. Interestingly, the coders and myself noted anecdotally that cluster labels on which they did not agree tended to have been labeled significantly differently, rather than being tangentially related. For example, a cluster that one coder had labelled "pursuing a diagnosis," the other coder had labelled "school experiences." However, in all of these cases, the coders noted that the content of the cluster in question had included content related to both themes each

20

one of them had identified, and both coders had struggled to synthesize the content of the cluster into a single theme.

In other words, the coders were in consensus that a given cluster contained each other's chosen theme, but had come to a different conclusion regarding which theme was more salient or overarching. When reconciling these coding conflicts, the coders and I prioritized themes based on two metrics, (a) whether one of the themes was novel and (b) which theme was perceived to be more specific. In the example given above, we reached the consensus to label the cluster "pursuing a diagnosis" because another cluster related to "school experiences" had already been identified.

Next, the results of labelling the TopEx scatter plot with identified themes were examined. I then made reasonable comparisons between the clusters produced by TopEx and the diagnostic criteria in the DSM-5. Specifically, I use semantic interpretations of the meaning communicated by the clusters to discuss the extent to which each DSM-5 criterion is represented by the participants' concerns.

In addition to the DSM-5, I also compared the results produced by TopEx to the chaptered contents of Barkley's (2018) widely well-regarded handbook, *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment* (*Fourth Edition*). This particular handbook was selected based on consultation with experts in the field of ADHD who cite Barkley's (2018) work as well-respected and a frequent choice to instruct clinicians in the diagnosis and treatment of ADHD. Furthermore, this particular handbook addresses both childhood and adulthood presentations of ADHD, which is an appropriate fit for a dataset which is expected to contain reflections of concerns across the lifespan.

Finally, in addition to making comparisons to the above clinical sources, I also made reasonable comparisons between the clusters produced by TopEx and the common themes identified in previous qualitative studies. Regarding the coding of the data and interpretation of the results, it is important to acknowledge the positionality of myself (the researcher) and the two coders who worked on this project.

The researcher and coders acknowledge relevant biases and perspectives they hold which may influence their presentation of these results. The researcher is a practicing clinician in training who seeks to specialize in working with neurodivergent adults; they hold great fondness for their neurodivergent clients and assume a strengths-based approach in their clinical work. This strengths-based approach includes humanistic influences. Thus, the researcher has a bias against deficit-based conceptualizations of ADHD. Furthermore, the researcher considers intersectional feminist theory—and, by extension, feminist therapy—major influences on their work. They believe that centering issues of social justice is critical to delivering the highest-quality services to clients. The first coder is entering their doctoral studies in clinical psychology and has a background in the humanities. They do not hold direct biases in their research on ADHD; however, indirect influences on their research are orientations toward social constructivist theories, social justice, and dimensional models of psychopathology. Additionally, their background in textual analysis may have influenced their perception and interpretation of the text studied in this research. Holding a bachelor's in Psychology, the second coder has great interest in mental health as well as empathy towards those struggling. Additionally, she believes that those identifying with ADHD are overlooked by mental health professionals. Therefore, she has a preconceived idea of the experiences of those identifying with ADHD. By naming their relevant biases and perspectives before initiating the review process and remaining cognizant of

their beliefs throughout the review, the researcher and coders hope to minimize the influence of their biases on the results.

<div align="center">**Results**</div>

Identified Themes

For each subsample, TopEx produced a scatterplot of data clusters and a word cloud for each respective cluster. An example scatterplot (figure 1) and word cloud (figure 2) are provided below.
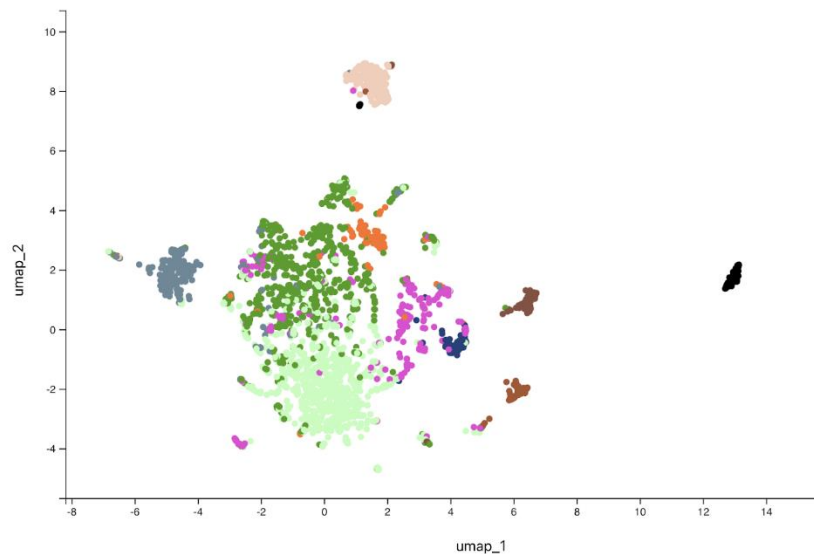

Figure 1. A scatterplot produced by TopEx modeling a subsample of the dataset
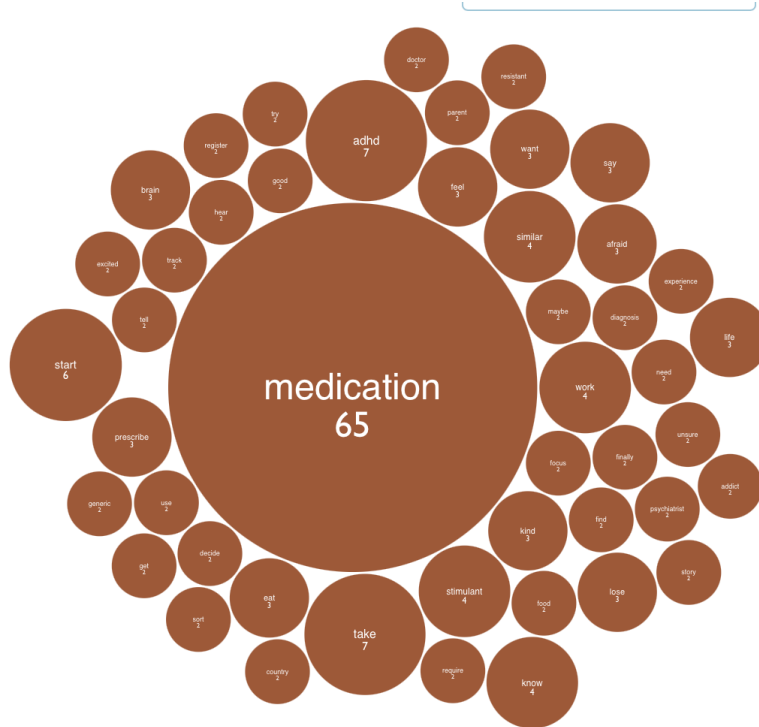
Figure 2. An example wordcloud produced by TopEx from the same subsample displayed in Figure 1

Across the 5 subsamples, 23 unique themes were identified by the coders. The themes are as follows: issues with productivity, issues with time management, difficulties with attention, difficulty completing tasks, comprehension issues, difficulties with reading habits, lack of motivation, feeling misunderstood by others, interpersonal distress, negative self-concept/low self-esteem, lack of social support, physiological symptoms, concerns with food and diet difficulty in school, employment concerns, side-effects of medication, positive effects of medication, lack of access to medication, negative experiences with health professionals, lack of professional support, seeking advice on getting diagnosed, facing stigma, and feeling failed by educational systems.

"Issues with productivity" reflected concerns individuals had with being able to complete their work (academic or employment) as well as not being able to complete enjoyable/recreational tasks related to hobbies; the context of "issues with time management" was similar. "Difficulties with attention" was characterized by individuals endorsing a struggle with being able to focus, once again both academically/professionally and recreationally. "Difficulty completing tasks," similar to aforementioned themes, related to both academic/professional and recreational situations. "Comprehension issues" related to individuals struggling to understand the communications made by others, both regarding interpersonal concerns or task-related instructions. This theme was distinct from aforementioned "feeling misunderstood by others" in that it was more focused on content-related misunderstandings as opposed to broader misunderstandings related to identity or personhood. "Difficulties with reading habits" focused on frustrations and struggles with reading particularly in the context of being unable to maintain attention or stamina while reading; commonly, posters contrasted their current difficulties with having been avid readers as children. "Lack of motivation" related to struggling to develop and maintain motivation for tasks in multiple domains of life.

"Feeling misunderstood by others" was characterized by interpersonal difficulties in terms of perceiving oneself as being judged by others, rejected for being different, or having their intentions misinterpreted. This included concerns regarding being labeled as "lazy" or a troublemaker by teachers, parents, peers, and romantic partners. "Interpersonal distress" covered relationship conflicts and stressors across multiple relationship types, including parental, romantic, and platonic friendship. Stressors were also diverse but notably included conflicts that posters perceived as being caused in part by symptoms of ADHD, such as their attention difficulties being a source of criticism or hurt feelings or having concerns with emotional

25

sensitivity. "Negative self-concept/low self-esteem" heavily featured endorsements of shame and self-blame for the difficulties people faced, as well as internalized beliefs regarding interpersonal deficits. "Lack of social support" included concerns regarding social isolation as well as having a social network but not receiving helpful support from those people.

"Physiological symptoms" included individuals discussing complaints such as increased body tension and gastrointestinal problems. "Concerns with food and diet" included concerns such as gastrointestinal issues (e.g., food allergies/sensitivities, frequently upset stomach), forgetting to eat, and disordered eating behaviors. "Difficulty in school" included individuals currently struggling in an educational system and individuals reflecting on their personal histories. "Employment concerns" was similar to "difficulty in school" in that it related to past and present concerns; it included issues with struggling to find and maintain appropriate employment.

"Side-effects of medication" featured content related to individuals discussing current or prospective medications and their experienced side-effects, as well as how to cope with these side-effects. "Positive effects of medication" focused on discussions where posters celebrated or expressed relief regarding the positive effects medication has had on their lives. "Lack of access to medication" was focused on individuals discussing barriers towards getting desired medication prescribed, including due to a formal diagnosis being required and perceiving medical providers as stigmatizing them as "drug-seeking."

"Negative experiences with health professionals" encompassed a variety of interaction types with a variety of professions, including general practitioners, psychiatrists, and therapists; these negative experiences often included an element of feeling invalidated, labelled as drug-seeking, or believing one was misdiagnosed. "Lack of professional support" included topics such

26

as barriers to accessing therapy services or believing the supports they were receiving to be inadequate. "Advice on getting diagnosed" captured both individuals asking others for advice on how to pursue a formal diagnosis and people sharing their own experiences with the diagnostic process to help others.

"Facing stigma" processed concerns related to being stigmatized in a variety of settings as well as internalized stigma. Finally, "feeling failed by educational systems" was characterized by posters reflecting on ways in which they believed their educational institutions to have failed them in the past, such as by failing to provide accommodations adequately or having their struggles overlooked.

Comparisons to *DSM-5*

Several themes identified in the dataset can be reasonably connected to the criteria and affiliated features of ADHD present in the *DSM-5*. Notably, issues with productivity, issues with time management, difficulties with attention, difficulty completing tasks, lack of motivation, and difficulties with reading habits are related to core features of ADHD. "Issues with productivity" and "difficulty completing tasks" are related to criterion 1d, "often does not follow though on instructions and fails to finish schoolwork, chores, or duties in the workplace." "Issues with time management" relate to criterion 1e, "often has difficulty organizing tasks and activities…e.g., has poor time management, fails to meet deadlines." "Lack of motivation" is related to criterion 1f, "often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort." Finally, "difficulties with attention" and "difficulties with reading habits" are related to criterion 1b, "often has difficulty sustaining attention in tasks or play activities (e.g. … lengthy reading)." Notably, these themes are all connected to the inattentive criteria. While less direct connections may be drawn between various themes and the hyperactivity criteria (e.g., one may hypothesize

that the criterion of "often interrupts or intrudes on others" may contribute to themes such as "difficulty in school" and "interpersonal distress"), these connections are less overt.

Many other themes identified in this study connect to the affiliated features of ADHD described in the *DSM-5*. The *DSM-5* reports "delays in language, motor, or social development," often co-occur, which connects to themes such as "feeling misunderstood by others," "comprehension issues," "lack of social support," and "interpersonal distress." As covered by themes related to "difficulty in school" and "employment concerns," affiliated issues also cover impairments in academics and the workplace. Further emphasizing connection to interpersonal-related themes, the *DSM-5* further references interpersonal relationship distress in a section describing Functional Consequences of the disorder, noting "peer neglect … and rejection," "elevated interpersonal conflict," and "lower self-esteem relative to peers." This last quotation relates to the identified theme, "negative self-concept/low self-esteem."

All in all, 13 of the 23 themes identified by this study can be reasonably compared to content present in the *DSM-5* regarding ADHD. Notably, the majority of these themes are related to inattentive symptom criteria, affiliated features, or functional consequences. The 10 themes which are not reasonably considered captured by the *DSM-5* focused on a few broader types of concern, such as medication-related issues ("side-effects of medication," "lack of access to medication," and "positive effects of medication"), interactions with societal systems ("seeking advice on getting diagnosed," "negative experiences with health professionals," "lack of professional support," "facing stigma," and "feeling failed by educational systems"), and physical health concerns ("physiological symptoms" and "concerns with food and diet").

Comparisons to a Treatment Manual

Comparisons will now be made between the dataset and Barkley's (2014) seminal

treatment manual addressing the treatment of ADHD. This text is comprised of four main parts:

(a) the nature of ADHD, (b) assessment of ADHD, (c) treatment of children and adolescents, and

(d) treatment of adults (Barkley, 2014). Because this study is primarily concerned with the

experience of adults, analysis will focus on parts 1 and 4, though part 3 will be alluded to when

indicated. Many of the themes identified in the present study are captured by the treatment

manual in multiple sections (e.g., discussed in part 1 while establishing symptomatology and

again in part 4 while addressing treatment strategies). For the sake of avoiding redundancy, I will

avoid overly reiterating the presence of a theme once it has already been compared to the text

once.

In part 1, Barkley (2014) first discusses the history and *DSM-5* diagnostic criteria of

ADHD; therefore, all of the themes of the study which were already compared to the *DSM-5*

core criteria are also captured in this first part of the treatment manual. Interestingly, Barkley

(2014) also contains a section in part 1 dedicated to emotion dysregulation, emphasizing that this

is a "core component" of ADHD. Our coders did not label a theme related to emotion

dysregulation concerns; however, both coders observed that concerns regarding emotion

regulation were often present within other themes, such as "interpersonal distress" and "negative

self-concept/low self-esteem."

Furthermore, part 1 of the treatment manual also includes chapters on educational and

occupational impairments, which are captured by this study's themes of "difficulty in school"

and "employment concerns." Several themes regarding interpersonal relationships,

"comprehension issues," "lack of social support," and "interpersonal distress" are captured in the

treatment manual's sections on childhood and adult relationships (Barkley, 2014). However,

another interpersonal theme, "feeling misunderstood by others" did not as reasonably compare to the content of these sections. Lastly regarding part 1, there is also a section which reviews health problems commonly faced by people identifying with ADHD (Barkley, 2014). Themes which compare to this section are "physiological symptoms" and "concerns with food and diet."

Regarding themes more focused on childhood, "feeling failed by educational systems" is captured by a section in part 3 which discusses the treatment of ADHD in school settings (Barkley, 2014). In this section, Barkley (2014) identifies common ways in which traditional school systems struggle to provide affirming care and supports for ADHD students and acknowledges the negative impact this can have on the children. In part 4, Barkley (2014) elaborates on how ADHD impacts people in adulthood and these impacts can be compared to identified themes such as "feeling misunderstood by others" and "facing stigma." The text also captures aspects of "negative experiences with health professionals" and "lack of professional support" by discussing how counselors can learn to be more effective and avoid misdiagnosis (Barkley, 2014). The themes "side-effects of medication," "lack of access to medication," and "positive effects of medication" are also all reasonably compared to a section in part 4 regarding pharmacology (Barkley, 2014).

Ultimately, only 1 of 23 themes cannot be reasonably compared to content in Barkley's (2014) treatment manual: "seeking advice on getting diagnosed." This theme was notably the only theme present that was explicitly focused on getting advice from peers, rather that being a broader theme about discussing experiences that included *some* posts which were advice-seeking.

Discussion

30

I hypothesized that the following topics would feature prominently in the results: (a) experiences of social rejection, (b) medication concerns, including struggling with gaining access to medication, (c) barriers to obtaining a clinical diagnosis, (d) seeking advice on coping strategies, and (e) celebrations of successes and/or uplifting positive aspects of having ADHD.

My first hypothesis regarding social rejection was supported by the following identified themes: "feeling misunderstood by others," "lack of social support," "interpersonal distress," and "facing stigma." These themes all contained content related to social rejection, distress, and isolation. All of these themes appeared across subsamples with significant redundancy, indicating a high level of confidence in these themes. These results are consistent with previous research, suggesting that interpersonal distress is widespread and concerning for people identifying with ADHD (Canu & Carlson, 2007); these concerns may be particularly salient for people identifying with ADHD because research suggests that many people with ADHD have heightened sensitivity to rejection (Babinksi et al., 2019; Bondu & Esser, 2015), so it is logical that several distinct themes related to interpersonal difficulties would appear in the sample. Given that lack of social support is significantly correlated with lower quality of life (Turner & Brown, 2010), the consistent appearance of these difficulties in the literature is concerning.

My second hypothesis, that medication concerns would be featured heavily in the data, was similarly well-supported by the themes "side-effects of medication," "lack of access to medication," and "positive effects of medication." Notably, themes related to medication occurred at a particularly high rate across subsamples. Most experts still regard a mix of pharmacology and psychotherapy as the gold standard treatment for ADHD (e.g., Manos & Short, 2023; Barkley, 2003), and topics related to medication, including its benefits, appear to be similarly salient to this sample. Interestingly, while previous qualitative studies made note of a

focus on receiving formal supports such as therapy services (e.g, Stenner, O'Dell, & Davies, 2018; Osagiede-Omon, 2020) medication specifically was not highlighted as its own theme with frequency. Thus, identifying three distinct themes related to pharmacological treatment specifically is a comparatively unique finding. This may be due to a variety of factors, such as the differences in setting. For example, participants also expressed concerns with being stigmatized and perceived as drug-seeking, which could indicate that individuals felt more comfortable discussing their desires for medication in an anonymous forum by and for peers.

The third hypothesis, regarding barriers to obtaining clinical diagnosis, was partially supported by the theme "seeking advice on getting diagnosed" which appears to indicate that many posters had concerns with navigating the process. The themes "lack of access to medication" and "lack of professional support" also contained content that touched on these concerns, though less directly. "Lack of access to medication" and "lack of professional support" occurred with some redundancy within the sample. For example, many posts clustered under the theme "lack of access to medication" involved identifying the barrier of needing a formal diagnosis before medication would be prescribed and struggling to obtain that documentation. These findings are consistent with the existing literature, where participants across multiple studies expressed issues with receiving adequate support (Osagiede-Omon, 2020; Waite & Tran, 2010; Zwennes & Loth, 2019).

The fourth and fifth hypotheses were not supported by the results. While many themes were identified related to having skills deficits, such as "issues with time management" and "difficulties with reading habits," these clusters did not specifically focus on advice-seeking related to these topics. This may be related to the nature of how TopEx operates as a natural language processor; possibly, advice-seeking was a more diffuse concept that had less

quantitative semantic weight compared to "most informative phrases" which related to the symptoms or struggles themselves. Alternatively, perhaps the members of this subreddit were more inclined to use the forum as a means of venting or catharsis than a source of advice-seeking. While advice-seeking was also not a prominent theme in past research, participants in other studies did emphasize various coping skills they had learned (Gully, 2009; Jenry & Jones, 2011; Schaedler, 1999; Waite & Tran, 2010). Finally, the lack of support with these themes could also be a result of the choice to only analyze the initial post contents, as opposed to also including replies to the initial posts. It is likely that many of the responses to the initial posts would have included more advice-giving and success stories in response an original poster's distress.

Relatedly, all but two of the themes ("seeking advice on getting diagnosed" and "positive effects of medication") can be considered negatively valanced in that they focus on the issues and difficulties posters face(d) rather than successes. It is important to note, however, that this does not mean that no positive or celebratory posts existed in the sample, but rather that when coding the data, the coders determined that these types of posts were not in sufficient numbers to reasonably discount the frequency of which posts focused on struggles, difficulties, etc. Perhaps this result, which is a departure from previous studies where positively-valenced themes were often highlighted (e.g., Henry & Jones, 2011; Osagiede-Omon, 2020), is partially due to differences between what metrics (i.e., positivity versus negativity) would be most salient to a human coder versus a machine-learning tool. These results may also be partially owed to the framing of the subreddit community itself, which states, that r/ADHD is "an inclusive, disability-oriented peer support group for people with ADHD." Given that the community is described as being for disability-oriented support, it is reasonable to posit that posters may have a selection

bias towards using this space to focus on the disabling aspects of their diagnosis rather than their successes or any positive aspects of having ADHD. An interesting direction for future work would be to compare the cluster analysis of r/ADHD to another online peer community, such as r/ADHDers, which describes itself in a different framework ("ADHDers, unite! This subreddit was created to provide a positive, safe, and inclusive environment for those with ADHD to converse. We are proud members of the Neurodiversity Movement, which is also a part of Disability Rights Movement") to determine if different themes emerge.

Overall, the level of alignment between the themes identified from the Reddit sample was moderate for the *DSM-5* and very high for Barkley's (2014) treatment manual. Because the *DSM-5* is inherently more limited in focus due to being a diagnostic manual, these findings are logical. It did appear that the *DSM-5* captured a reasonable proportion of themes that were related to symptomatology and functional concerns; however, many of these themes were represented in the "additional features" section on ADHD rather than by the core criteria. When observing this point, it is important to remember that core diagnostic criteria are constructed with a focus on the features that are most unique to and best distinguish a group (i.e., people with ADHD) from others. Concerns such as interpersonal distress were clearly salient to people with ADHD are mentioned, but are not unique to the disorder, as interpersonal distress can be a feature caused or exacerbated by many mental health concerns.

It is encouraging for the state of gold standard ADHD treatment that all but one of the twenty-three themes identified by this study were present in the reviewed treatment manual. At the same time, this raises concerns with the fact that, over one decade after Barkley's (2014) manual was published, several themes were still present regarding the shortcomings of professional care, such as "negative experiences with health professionals," "lack of professional

support," and "facing stigma." This statement is made with the caveats that those themes included a variety of health professionals, not just therapists, and that the majority of the identified themes overall were negatively valanced.

Indeed, as discussed above, the subject matter as a whole was focused on difficulties, negative impacts, and concerns rather than positive or celebratory topics. These results may serve as an important acknowledgement that many people still struggle to access services and to feel affirmed by the services they do receive. At the same time, it is encouraging that many of the concerns people identifying with ADHD in this sample expressed are addressed in empirically supported treatments in the clinical literature rather than not being explicitly covered.

This study used innovative data sources (i.e., social media) and data analysis techniques (i.e., TopEx). The two coders and myself reflected on the experience of using these innovative methods, including positives and limiting factors to consider. Notably, our data was collected shortly before Reddit made major changes to its policies regarding its application programming interface (API); these changes severely limit the quantity and speed at which data scraping from Reddit can be performed, unless users pay substantial fees for the rights to engage with Reddit's API. Under this new policy, I likely would not have been able to conduct my dissertation without pursuing grant funding; this is a barrier towards replicating the study in the future. Despite these new difficulties, there were several benefits to using big data from social media; for example, the current sample represents a very large group of individuals who self-identify as having ADHD from presumably diverse backgrounds, allowing for a dataset that is much larger than other conventionally gathered research samples (Law et al., 1998). Additionally, this dataset represented participants expressing themselves without the influence of an interviewer's

presence and provided access to a mixture of historical and present-day concerns for the participants.

Both coders agreed that the ability to use TopEx as a clustering tool for initial data analysis was useful and interesting; they observed that the process was significantly less overwhelming and significantly faster than traditional qualitative coding, which was particularly important given the high volume of data they were asked to code. The coders noted that one barrier was determining the appropriate number of clusters for a subsample, because for a given cluster number (e.g., if the cluster number was set to fifteen), there would be both clusters that appeared redundant with each other (e.g., two clusters regarding medication in the same subsample) *and* clusters that appeared overly large/noisy. Therefore, it was challenging for the coders to determine if the number of clusters was appropriate. The coders also noted that sometimes a cluster appeared to be overly noisy or disparate in terms of theme, but as more sample phrases were examined, a theme that was "surprisingly conceptual" came to light. In other words, TopEx seemed to identify a meaningful cluster that would have most likely been overlooked by a human qualitative analyst imposing preexisting expectations. While a detailed analysis of this point is beyond the scope of this dissertation, it has interesting implications for the conceptualization of natural language processors and their interactions with semantic content that would be beneficial to explore in future works.

Regarding the topic of positionality and bias, both coders identified perspectives they held which may have biased their coding process. The first coder named an orientation towards social constructivism, social justice, and dimensionality models of psychopathology. The second coder named having significant empathy towards people with ADHD and a belief that this population is often overlooked by mental health professionals. Thus, the coders had to be

mindful of how their perspectives may have biased their interpretation of TopEx's clusters. It is notable that while they endorsed different backgrounds and perspectives, the coders had a relatively high (average 89.4%) level of agreement after the initial independent coding phase.

I stated that as the researcher, I held a bias towards a humanistic, strengths-based approach towards ADHD conceptualization and treatment. This approach influenced two of my hypotheses in particular: that there would be topics regarding seeking advice/support and regarding positivity/successes. Interestingly, these two hypotheses were also the only two to lack significant support by the results. As the primary responsible party for ultimately interpreting and presenting the results, I needed to remain mindful of how my desire to see my hypotheses supported, particularly the ones most informed by my clinical worldview, may influence how themes were ultimately described and presented in this report.

Limitations

This study presents several limitations. First, due to the nature of social media and how the data was collected, the participants are unidentifiable in terms of demographics, including whether or not they have been formally diagnosed with ADHD. There may also be issues with the data being skewed by repeat posters. As mentioned above, another concern with using Reddit as the data source is that Reddit's new API policies make it less accessible for researchers to use big data from the social media platform as I did in this study; this creates concerns with the study being more difficult to replicate. However, if funding were available, replication attempts would still be possible and valuable.

While TopEx has many advantages, it also has its own limitations. As noted above while discussing the experience of the two coders, determining the appropriate number of clusters to

impose was difficult at times and ultimately subjective. This introduces a potential source of bias into the results depending on how a coder chooses to parse the data. Furthermore, while using machine-learning algorithms can reduce certain types of bias, it also introduces its own forms of bias and fails to address others. The theme labels were still ultimately determined by myself and human coders who imposed a level of meaning and importance on the clusters' contents. Machine-learning algorithms are also vulnerable to being biased by the corpus that trains them; because TopEx learns in a bottom-up fashion (Olex et al., 2020) the associations it builds between words and semantic connections it makes are dependent on the way language is used in the training corpus. Thus any human biases that are reflected in the training corpus, such as general societal stigma against mental health conditions (Godfrey et al., 2021), will likely also emerge in the results.

Future Directions and Clinical Implications

It is encouraging that themes identified in this study had a high level of agreement with one gold standard treatment manual (Barkley, 2014) for ADHD. This indicates that clinicians who seek out specialized training/education are more likely to be prepared with treatments that align with the concerns of adults with ADHD. Clinicians may want to consider the emphasis placed by participants in this study on the impact their ADHD has on interpersonal relationships and educational/employment settings. They may also want to consider systemic barriers that their clients have faced in the past and continue to face in the present, and work with their clients to understand how those barriers impact their ability to address their treatment concerns.

Despite the high level of agreement between the identified themes and Barkley's (2014) manual, it is notable that several themes related to dissatisfaction with treatment providers and access to formal supports also emerged. This indicates that there are still gaps between what

38

services/supports people identifying with ADHD want and the services they perceive themselves as having received or had the opportunity to receive. Therefore, future directions may include (a) more focused research to understand themes that emerge regarding people identifying with ADHD's experiences in different treatment settings to better understand these gaps, (b) reducing systemic barriers towards accessing services, such as advocating for societal changes that will make treatments fair and affordable, and (c) furthering education for healthcare professionals across disciplines regarding how to provide effective treatments to individuals with ADHD. Related to this final point, there was a particular focus by participants on issues regarding medication, which is considered part of the gold standard ADHD treatment (Manos & Short, 2023). Clinical implications thus include considering how people identifying with ADHD who are interested in seeking medication to help treat their symptoms can be supported in understanding their options and gaining access to appropriate pharmacological treatment without stigmatization.

Other avenues for future research would include using TopEx to attempt to replicate the present study and/or to analyze related datasets. For example, the themes identified within r/ADHD may be distinct from themes which would be identified in related but distinct subreddits such as the aforementioned r/ADHDers, which describes its community as having a stronger activist component, or more specialized subreddits such as r/TwoXADHD, a subreddit specifically for women and people assigned-female-at-birth with ADHD. Machine-learning data analytic techniques could also be expanded to other social media platforms, such as X or Tumblr, which also have notable self-advocacy communities. Additionally, researchers may consider using TopEx or other machine-learning tools to analyze more traditional qualitative sources, such as semi-structured interviews with participants.

Conclusions

Using TopEx to analyze a sample of social media posts made by individuals who identify as having ADHD, this study identified twenty-three unique themes which addressed a variety of topics including primarily inattentive-related symptoms, interpersonal-related concerns, topics related to medication, and topics related to broader systems and formal supports. Surprisingly, few themes related to advice-seeking or positive/uplifting topics were identified, with the majority of the topics being negatively-valanced. However, there was moderate agreement between the identified themes and the *DSM-5* and high agreement between the identified themes and the gold standard treatment manual (Barkley, 2014).

References

ADHD-alien. (2020, June 5). ADHD & Relationships [Tumblr post]. Retrieved from

      https://adhd-alien.tumblr.com/post/620118857621897216/how-adhd-can-influence-

      relationships-romantically

ADHD-alien. (2020, October 16). ADHD & working from home [Tumblr post]. Retrieved from

      https://adhd-alien.tumblr.com/post/632148548460429313/adhd-can-feel-like-it-gets-

      worse-when-we-get

Ablednt. (2020, May 31). Hey so I've seen it happening a lot… [Tumblr post]. Retrieved from

      https://ablednt.tumblr.com/post/619668145064509440/hey-so-ive-seen-it-happening-a-

      lot-and-am-by-far

Adhdbreadbin. (2020, October 11). Been thinking about how people say adhd manifests…

      [Tumblr post]. Retrieved from

      https://adhdbreadbin.tumblr.com/post/631713501454483456/been-thinking-about-how-

      people-say-adhd-manifests

Alienjuuce. (2021, March 27). Y'know what ruffles my feathers? [Tumblr post]. Retrieved from

      https://alienjuuce.tumblr.com/post/646802082275491840/yknow-what-ruffles-my-

      feathers-the-diagnostic

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*

      (5th ed.). Arlington, VA: Author.

Babinski, D. E., Kujawa, A., Kessel, E. M., Arfer, K. B., & Klein, D. N. (2019). Sensitivity to

      peer feedback in young adolescents with symptoms of ADHD: examination of

neurophysiological and self-report measures. *Journal of abnormal child psychology*, *47*(4), 605-617.

Badger [@voxeldogs]. (2021, October 27). *#actuallyADHD #ADHD tip of the day: buy these. Keep them on/in your nightstand. Get to bed and forget to brush* [Tweet]. Twitter. https://twitter.com/voxeldogs/status/1453223280840749061

Barkley, R. A. (Ed.). (2014). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment.* Guilford Publications.

Biggest-gaudiest-patronuses. (2021, February). Tips for dealing with sensory issues… [Tumblr post]. Retrieved from https://biggest-gaudiest-patronuses.tumblr.com/post/641773359422734336/tips-for-dealing-with-sensory-issues-when-youre

Biggest-gaudiest-patronuses. (2021, July). It's just a small thing, but breaking down tasks… [Tumblr post]. Retrieved from https://biggest-gaudiest-patronuses.tumblr.com/post/656534023505657857/its-just-a-small-thing-but-breaking-down-tasks

Bondü, R., & Esser, G. (2015). Justice and rejection sensitivity in children and adolescents with ADHD symptoms. *European child & adolescent psychiatry*, *24*(2), 185-198.

Dumit, J. (2006). *Illnesses you have to fight to get: Facts as forces in uncertain, emergent illnesses.* Social Science and Medicine, 62(3), 577–590.

Canu, W. H., & Carlson, C. L. (2007). Rejection sensitivity and social outcomes of young adult men with ADHD. *Journal of Attention Disorders*, *10*(3), 261-275.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: a black feminist critique
of antidiscrimination doctrine, feminist theory and antiracist politics". *University of
Chicago Legal Forum.1989* (1): 139–167. ISSN 0892-5593. Archived from the
original on 15 August 2017.

Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S.
Dennis & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 167-184).
Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating
qualitative research. *Australian & New Zealand Journal of Psychiatry*, *36*(6), 717-732.

Godfrey, E., Fuermaier, A. B., Tucha, L., Butzbach, M., Weisbrod, M., Aschenbrenner, S., &
Tucha, O. (2021). Public perceptions of adult ADHD: Indications of stigma?. *Journal of
Neural Transmission, 128*(7), 993-1008.

Gully, D. (2009). *A study of the talent development of gifted individuals with attention deficit
hyperactivity disorder*. The College of William and Mary.

Harborne, A., Wolpert, M., & Clare, L. (2004). Making sense of ADHD: a battle for
understanding? Parents' views of their children being diagnosed with ADHD. *Clinical
Child Psychology and Psychiatry, 9*(3), 327-339.

Hare, E. J. (2015). *Talking about race: How do White clinicians engage in dialogue about race
in cross-racial therapy with Black clients?* Smith College.

Henry, E., & Hill Jones, S. (2011). Experiences of older adult women diagnosed with attention
deficit hyperactivity disorder. *Journal of women & aging*, *23*(3), 246-262.

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., Greenhill, L.L., Howes, M.J., Secnik, K. & Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. *American Journal of Psychiatry*, 163(4), 716-723.

Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (1st ed., pp. 3-34). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259-284.

Law, M., Stewart, D., Letts, L., Pollock, N., Bosch, J., & Westmorland, M. (1998). Guidelines for critical review of qualitative studies. *McMaster University occupational therapy evidence-based practice research Group*, 1-9.

Manos, M. J., & Short, E. J. (2023). A new paradigm for adult ADHD: A focused strategy to monitor treatment. *Cleveland Clinic journal of medicine, 90*(7), 413-421.

Michielsen, M., de Kruif, J. T. C., Comijs, H. C., van Mierlo, S., Semeijn, E. J., Beekman, A. T., & Kooij, J. S. (2018). The burden of ADHD in older adults: A qualitative study. *Journal of attention disorders*, *22*(6), 591-600.

Morris, E. W., & Perry, B. L. (2017). Girls behaving badly? Race, gender, and subjective evaluation in the discipline of African American girls. *Sociology of education*, *90*(2), 127-148.

Olex, A. L., DiazGranados, D., McInnes, B. T., & Goldberg, S. (2020). Local Topic Mining for
Reflective Medical Writing. AMIA Summits on Translational Science Proceedings, 2020,
459.

Olex, A. L., French, E., Burdette, P., Sagiraju, S., Neumann, T., Gal, T. S., & McInnes, B. T.
TopEx: Topic Exploration of COVID-19 Corpora. Results from the Biocreative VII
Challenge Track 4.

Ottosen, C., Larsen, J. T., Faraone, S. V., Chen, Q., Hartman, C., Larsson, H., Petersen, L., &
Dalsgaard, S. (2019). Sex differences in comorbidity patterns of attention-
deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent
Psychiatry, 58*(4), 412-422.

Osagiede-Omon, F. (2020). *A bioecological study of twice-exceptional African American males:
The intersection of gift, disability, race, and gender* (Doctoral dissertation, University of
Southern California).

Painter, C. A., Prevatt, F., & Welles, T. (2008). Career beliefs and job satisfaction in adults with
symptoms of attention-deficit/hyperactivity disorder. *Journal of Employment
Counseling*, *45*(4), 178-188.

Peterson, B. S., Trampush, J., Brown, M., Maglione, M., Bolshakova, M., Rozelle, M., Miles, J.,
Pakdaman, S., Yagyu, S., Motala, A. & Hempel, S. (2024). Tools for the Diagnosis of
ADHD in Children and Adolescents: A Systematic Review. *Pediatrics,* e2024065854.

[Potential-Search-376]. (2022, August, 17). *Does ADHD sometimes make it harder to have sex?*
[Online forum post]. Reddit.

https://www.reddit.com/r/ADHD/comments/wqiq8l/does_adhd_sometimes_make_it_har

der_to_have_sex/

[r/ADHD]. (2008, October 28). *ADHD.* [Online forum]. Reddit.

https://www.reddit.com/r/ADHD/

[r/ADHDers]. *ADHDers.* [Online forum]. Reddit. https://www/reddit/com/r/ADHDers/

[r/TwoXADHD]/ *TwoXADHD.* [Online forum]. Reddit. https://reddit.com/r/TwoXADHD/

Sattelberg, W. (2021, July 11). *The Demographics of Reddit: Who Uses the Site?* Alphr.

https://www.alphr.com/demographics-reddit/

Schaedler, J. S. (1999). *A phenomenological study of the developmental course and adult*

outcome of women with attention-deficit/hyperactivity disorder. United States

International University.

Sloan, D., & Marx, B. (2004). Taking pen to hand: Evaluating theories underlying the written

disclosure paradigm. *Clinical Psychology: Science and Practice, 11*(2), 121-137.

So1987. (2021, June 17). To the kids who spent hours hyperfixating… [Tumblr post]. Retrieved

from https://so1987.tumblr.com/post/654280839249674240/to-the-kids-who-spent-hours-

hyperfixating-on-and

Stenner, P., O'Dell, L., & Davies, A. (2019). Adult women and ADHD: On the temporal

dimensions of ADHD identities. *Journal for the Theory of Social Behaviour*, *49*(2), 179-

197.

Thevirgodoll. (2021, March 11). Coping Mechanisms Guide [Tumblr post]. Retrieved from

    https://thevirgodoll.tumblr.com/post/645388602987134976/what-are-healthy-coping-

    mechanism-youd-recommend

Turgay, A., Goodman, D. W., Asherson, P., Lasser, R. A., Babcock, T. F., Pucci, M. L., ... &

    ADHD Transition Phase Model Working Group. (2012). Lifespan persistence of ADHD:

    the life transition model and its application. *The Journal of clinical psychiatry, 73*(2),

    10337.

Turner, R. J., & Brown, R. L. (2010). Social support and mental health. *A handbook for the study

    of mental health: Social contexts, theories, and systems*, *2*, 200-212.

Waite, R., & Tran, M. (2010). ADHD among a cohort of ethnic minority women. *Women &

    health*, *50*(1), 71-87.

Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R., Loo,

    S.K., Carlson, C.L., McBurnett, K. & Lahey, B. B. (2012). Validity of DSM-IV attention

    deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of abnormal

    psychology, 121*(4), 991.

Williamson, D., & Johnston, C. (2015). Gender differences in adults with attention-

    deficit/hyperactivity disorder: A narrative review. Clinical Psychology Review, 40, 15–

    27.

Zwennes, C. T., & Loth, C. A. (2019). "Moments of failure": coping with Attention Deficit

    Hyperactivity Disorder, sleep deprivation, and being overweight: A qualitative

    hermeneutic–phenomenological investigation into participant perspectives. *Journal of

    addictions nursing*, *30*(3), 185-192.

Vita

Gabby Scalzo was born in Virginia Beach, Virginia and is an American citizen. They graduated from Ocean Lakes High School in Virginia Beach, VA in 2009, then received their Bachelor of Science in Psychology and Neuroscience from Virginia Polytechnic Institute and State University with Highest Honors, in Blacksburg, VA in 2013. In the fall of 2013, Gabby enrolled in the Clinical Psychology Doctoral Degree program at Virginia Commonwealth University in Richmond, VA under the mentorship of Drs. Scott Vrana and Jared Keeley. After graduation, Gabby plans to put their excellent training to use in private practice serving the Richmond area.