VALIDITY OF THE GLOBAL APPRAISAL OF INDIVIDUAL NEEDS SHORT SCREENER (GSS) IN AN INTEGRATED YOUTH HEALTH SERVICE

by

Raymond Khanano

B.Sc.Kin, McMaster University, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Population and Public Health)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2019

© Raymond Khanano, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

VALIDITY OF THE GLOBAL APPRAISAL OF INDIVIDUAL NEEDS SHORT SCREENER (GSS) IN AN INTEGRATED YOUTH HEALTH SERVICE

submitted by Raymond Khanano in partial fulfillment of the requirements for
the degree of Master of Science
in Population and Public Health

Examinig Committee:

Chris Richardson, Ph.D.
Supervisor
Skye Barbic, B.Sc.OT., M.Sc., Ph.D.
Supervisory Committee Member
Joseph Puyat, M.Sc., MA, Ph.D.
Supervisory Committee Member
Trevor Dummer, BA, M.Sc., Ph.D.
Additional Examiner

Additional Supervisory Committee Members:

Joanna Henderson, Ph.D.
Supervisory Committee Member
Abstract

OBJECTIVES:

There is increasing interest in the identification of mental illnesses among youth through routine screening in clinical settings. One tool currently being used in Canada is the Global Appraisal of Individual Needs (GAIN) Short Screener (GSS). The first objective of this study was to provide the results of a detailed literature search of existing literature on the GAIN and GSS. The second and third objectives were to contribute to literature on the validity of the GAIN by examining the sensitivity and specificity of the GSS and the GSS Internalizing Disorder Screener (GSS-I) respectively in an integrated youth health service.

METHOD(S):

Participants [n=201, 44% women, mean age 21.3 (SD=2.0) years] were recruited from an integrated youth health service in Vancouver, British Columbia. Data collection of the GSS, the Kessler Psychological Distress Scale 10 (K10), the Patient Health Questionnaire 9 (PHQ-9), and the Generalized Anxiety Disorder Scale 7 (GAD-7) was completed. Pearson’s correlational coefficient, Cronbach’s Alpha, Cohen’s Kappa coefficient, sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) of the GSS and GSS-I were calculated. Receiver Operator Characteristic (ROC) curves were generated, and Youden’s J Statistic was calculated to determine optimal cut-points for the tools in this study. Finally, Area Under the Curve (AUC) values for each ROC curve were calculated.

RESULTS:

Sensitivity and specificity of the GSS total score referenced with the K10 total score were 100% and 9.6% respectively. An optimal cut-point of 7 yielded sensitivity and specificity values...
of 96.0% and 40.4% respectively. A similar pattern of results was found for the GSS-I in relation to the K10, PHQ-9, and GAD-7.

CONCLUSION:

The results indicated that the GSS and GSS-I have acceptable sensitivity but poor specificity that could be improved via the use of an optimized cut-point. In the context of an integrated youth health service where follow-up assessments are readily available low specificity may be acceptable. Future research examining the validity of the GSS and GSS-I should examine the sensitivity and specificity of the tools with a clinical assessment gold standard, for use in integrated youth health services as well as other non-clinical settings.
Lay Summary

The Global Appraisal of Individual Needs Short Screener (GSS) is a brief, 23-item tool used by researchers and clinicians to screen for mental health, substance use, and crime and violence problems. This setting of this study was an integrated youth health service, where young people can access primary care, psychological care, and integrated case management all under one roof. The primary goal of this study was to examine the accuracy of the GSS in its ability to screen youth for these outcomes in the context of an integrated youth health service. This study demonstrated that there are public health benefits associated with mental health screening in an integrated youth health service; however, further evidence may be required to determine if the GSS functions effectively in this context. It is recommended that future studies examining the performance of this tool use a clinical assessment to capture its validity more accurately.
Preface

At the time of writing this thesis, I declare that the identification and design of the research program was initially determined by Dr. Skye Barbic, Dr. Chris Richardson, and myself. Original data collection was performed by Dr. Barbic and her team at Foundry. Analysis of the research data, reporting of the results, and creation of the tables and figures was performed by myself. I have written all other content presented in this thesis. Dr. Barbic, Dr. Henderson, Dr. Mathias, and Dr. Richardson were all involved in the editing and formatting of the writing contained in this thesis.

The Abstract, Chapters 2, 3, and 4 of this thesis are in the process of being submitted as original research for publication. As the primary author of this study I completed the majority of the manuscript writing, data analysis, and reporting on the findings. Dr. Barbic and her team at Foundry were responsible for the collection and management of data. All authors contributed to the editing and formatting of the manuscript.

In the primary study, eligible youth, aged 19 and over, followed consenting protocols outlined by the University of British Columbia Ethics Board (REB ID#H17-00127). For youth aged 18 or less, assent protocols, followed by a trained research team member, were used.
Table of Contents

Abstract .................................................................................................................................................. iii

Lay Summary ........................................................................................................................................ v

Preface .................................................................................................................................................. vi

Table of Contents ................................................................................................................................ vii

List of Tables ....................................................................................................................................... xi

List of Figures ....................................................................................................................................... xii

List of Abbreviations .......................................................................................................................... xiii

Acknowledgements .............................................................................................................................. xiv

Dedication ............................................................................................................................................. xv

Chapter 1: Background ....................................................................................................................... 1

1.1 Definitions of Youth and Emerging Adulthood ........................................................................... 1

1.2 Mental Health of Youth .................................................................................................................. 3

1.3 Classification and Diagnosis of Mental Illnesses ....................................................................... 4

1.3.1 The Diagnostic and Statistical Manual for Mental Disorders .............................................. 4

1.3.2 Trans-Diagnostic Symptom Clusters ...................................................................................... 6

1.3.3 Clinical Staging Model of Mental Illness .............................................................................. 7

1.4 Integrated Youth Health Centres ................................................................................................. 9

1.5 Mental Health Screening in Youth ............................................................................................... 10

1.5.1 Screening Youth in Schools ................................................................................................... 11

1.5.2 Screening Youth in Clinical Settings ..................................................................................... 12

1.6 The GAIN Family of Measures .................................................................................................. 14

1.6.1 Clinical Interviews, Assessments, and Screening Tools ......................................................... 14
1.6.2 The Global Appraisal of Individual Needs ................................................................. 15
1.6.3 The GAIN Short Screener (GSS) ........................................................................ 16

1.7 Psychometric Properties of Clinical Assessments and Screening Tools .............. 17
1.7.1 Validity .................................................................................................................. 17
1.7.2 Reliability ............................................................................................................. 18

1.8 Objectives and Study Hypotheses ........................................................................... 19

1.9 Literature Review of the GAIN and GSS ................................................................. 19
1.9.1 Literature Search Strategy .................................................................................... 19
1.9.2 Databases ............................................................................................................ 20
1.9.3 Article Screening .................................................................................................. 20
1.9.4 Article Categorizing .............................................................................................. 21

1.10 Literature Search Results ........................................................................................ 21
1.10.1 Studies Which Used the GAIN and GSS as Assessment or Screening Tools ...... 23
1.10.2 Studies Which Used the GAIN and GSS as Gold Standards (Reference Tools).. 28
1.10.3 Studies Examining the Validity of the GAIN and GSS ........................................ 31
1.10.3.1 GAIN Validity and Reliability ........................................................................ 31
1.10.3.2 GSS Validity .................................................................................................... 32
1.10.3.3 GSS Reliability ............................................................................................... 34
1.10.3.4 Other Studies Which Utilized or Studied the GAIN ......................................... 35

Chapter 2: Research Hypotheses and Methods ............................................................... 37

2.1 Research Objectives for the Analytical Studies ......................................................... 37
2.2 Study Setting and Recruitment ............................................................................... 37
2.2.1 Foundry ............................................................................................................... 37
2.3 Data Source .............................................................................................................. 38
2.4 Measures .................................................................................................................................................. 39
  2.4.1 GSS and GSS-I .................................................................................................................................. 39
  2.4.2 K10.................................................................................................................................................... 40
  2.4.3 PHQ-9 ................................................................................................................................................ 41
  2.4.4 GAD-7................................................................................................................................................. 42

2.5 Analyses ...................................................................................................................................................... 43
  2.5.1 Descriptive Statistics............................................................................................................................. 44
  2.5.2 Pearson’s Correlation Coefficient and Correlational Matrix ............................................................ 44
  2.5.3 Cronbach’s Alpha ............................................................................................................................... 45
  2.5.4 Cohen’s Kappa Statistic ...................................................................................................................... 45
  2.5.5 Sensitivity and Specificity ................................................................................................................... 46
  2.5.6 Positive and Negative Predictive Values ............................................................................................ 46
  2.5.7 ROC Curves ......................................................................................................................................... 47
  2.5.8 AUC .................................................................................................................................................... 48
  2.5.9 Missing Data Analysis ......................................................................................................................... 48

Chapter 3: Results ............................................................................................................................................ 50
  3.1 Descriptive Statistics................................................................................................................................. 50
  3.2 Bivariate Correlation Statistics ................................................................................................................ 52
  3.3 Cronbach’s Alpha ..................................................................................................................................... 54
  3.4 Cohen’s Kappa Coefficient ...................................................................................................................... 54
  3.5 Analysis of the GSS Total Disorder Screener Using the K10 as a Reference Tool .... 55
    3.5.1 GSS Sensitivity and Specificity ........................................................................................................... 55
    3.5.2 GSS PPV and NPV .............................................................................................................................. 56
    3.5.3 GSS ROC Curve ................................................................................................................................. 57
3.6 Analysis of the GSS-I using the K10, PHQ-9, and GAD-7 as Reference Tools

3.6.1 GSS-I Sensitivity and Specificity

3.6.2 GSS-I PPV and NPV

3.6.3 GSS-I ROC Curves

Chapter 4: Discussion

4.1 Summary of Findings

4.2 Over-Diagnosis and Over-Screening

4.3 Support for a Clinical Staging Model

4.4 Provincial and Pan-Canadian Recommendations for Standardized Use of the GSS

4.5 Future Directions of Research

4.6 Limitations

4.7 Applications of Research

Chapter 5: Conclusion

5.1 Overall Significance and Contributions of Research

Appendix A

Appendix B

Appendix C

Appendix D

Appendix E
List of Tables

Table 1 - Descriptive Statistics Results ........................................................................................................... 52
Table 2 - Pearson’s Correlational Matrix Results ................................................................................................ 53
Table 3 - GSS Cohen’s Kappa Results .................................................................................................................. 54
Table 4 - GSS-I Cohen’s Kappa Results ............................................................................................................... 55
Table 5 - K10, PHQ-9, and GAD-7 Cohen’s Kappa Results ............................................................................. 55
Table 6 - GSS Sensitivity and Specificity Results ............................................................................................... 56
Table 7 - GSS PPV and NPV Results .................................................................................................................. 56
Table 8 - GSS-I Sensitivity and Specificity Results ............................................................................................ 59
Table 9 - GSS-I PPV and NPV Results ............................................................................................................... 60
List of Figures

Figure 1 - Diagram Outlining Literature Search Process .......................................................... 22
Figure 2 - ROC Curve for the GSS and K10 ........................................................................... 57
Figure 3 - ROC Curve for the GSS-I and K10 ......................................................................... 61
Figure 4 - ROC Curve for the GSS-I and PHQ-9 .................................................................... 61
Figure 5 - ROC Curve for the GSS-I and GAD-7 .................................................................... 62
Figure 6 - GSS QQ Plot for Normality ....................................................................................... 127
Figure 7 - K10 QQ Plot for Normality ....................................................................................... 127
Figure 8 - PHQ-9 QQ Plot for Normality ................................................................................... 128
Figure 9 - GAD-7 QQ Plot for Normality .................................................................................. 128
Figure 10 - GSS and K10 Plot for Linearity .............................................................................. 129
Figure 11 - GSS-I and K10 Plot for Linearity .......................................................................... 129
Figure 12 - GSS-I and PHQ-9 Plot for Linearity ...................................................................... 130
Figure 13 - GSS-I and GAD-7 Plot for Linearity ..................................................................... 130
Figure 14 - GSS Questionnaire and Scoring Guide ................................................................. 132
Figure 15 - K10 Questionnaire and Scoring Guide ................................................................. 134
Figure 16 - PHQ-9 and GAD-7 Questionnaire and Scoring Guide ......................................... 137
List of Abbreviations

DSM-V – Diagnostic and Statistical Manual of Mental Disorders (Fifth Version)

GAD-7 – Generalized Anxiety Disorder Scale 7

GAIN – Global Appraisal of Individual Needs

GSS – Global Appraisal of Individual Needs Short Screener (Total Disorder Screener)

GSS-I – GSS Internalizing Problems Scale

ICD – International Classification of Diseases

K10 – Kessler Psychological Distress Scale 10

PHQ-9 – Patient Health Questionnaire 9

WHO – World Health Organization
Acknowledgements

I would like to extend the warmest thank you to my supervisor Chris Richardson. From the very first email I sent you requesting to be my supervisor, to all the phone call meetings we had in different time zones, you have been an incredibly supportive supervisor and wise teacher. I would also like to thank you for providing me with grant support in the form of a GRA. Thank you to Skye, Joanna, Joseph, Trevor, and Amy for taking time to be members of my thesis and examination committee. Your feedback and time are greatly appreciated and highly valued. I would like to extend a special thank you to Steven Mathias, who took time from his sabbatical to aid in the preparation of the manuscript of this study for publication.

Thank you to Shannon, Charlotte, and everyone over at CAMH for providing me with a novel challenge while tackling my thesis at the same time!

Thank you to Louise, Jacqueline, Nicole, Janae, and all my Soup Thursday pals at BCCHR. I had such a great time being part of the team during my time as a WorkLearn student.

Big shoutout to the SPPH homies, MPH study group friends, my fellow MSc cohort, and all my peers and colleagues I met through SPPH.

To Claudia and Johnny, I love you both. To Musks, I can’t wait until our next episode.

Finally, I would like to thank the everlasting love and support of my family. I love you so much Mom, Dad, Simon, Noora, Mary, Nansy, Matthew, Jasmine, Isabel, and all my aunts, uncles, and Nanny Setto.
To my late grandmother Liya Kiso,
& my niece Laya Noon Khanano.
Chapter 1: Background

The first chapter of this thesis provides relevant background information followed by a more focused literature review on core subject areas. The background begins by providing key definitions of terms that will be used throughout the study and highlighting why they are important to this study. The various terms assigned to young people such as “youth”, “adolescence”, and “young adult” will be contrasted, and this will give way to the discussion on mental illness in youth populations. A brief discussion of the classification of mental illnesses, encompassing traditional clinical techniques to more holistic and emerging approaches, are provided to describe the way in which symptoms and disruptions in behaviour emerge and become diagnosable mental disorders. The current study focuses on the use of a screening tool in an integrated youth health service, and all the background information presented will aid in building a knowledge base for the reader on this subject. Finally, this chapter will conclude with the results of a detailed literature search on the Global Appraisal of Individual Needs (GAIN) and its associated measures. This chapter will transition into the methodology and analytical plan that would arrive at the results of this study which examined the sensitivity and specificity of a screening tool used in an integrated youth health service.

1.1 Definitions of Youth and Emerging Adulthood

The United Nations Educational, Scientific and Cultural Organization (UNESCO) defines the period of “youth” as being between the ages of 15 and 24 years. In the literature there have also been distinctions made between the terms “adolescence” and “young adults” as developmental periods. For example, the Cleveland Clinic defines adolescence as a period of physical and social change in youth, marked by the onset of puberty, and having an ambiguous or societally-defined end period. The definition for young adult is also not explicitly defined,
and has been described similarly as a period of transition and identity exploration in early life. In both cases, the terms ascribed to young people are broad and reflect the societal changes that have contributed to prolonged periods of social development in youth. This developmental prolongation was explored by Arnett when he proposed the term “emerging adults” in 2000. He proposed a new developmental stage characterized by youth between the ages of 16 and 25 who are transitioning out of adolescence and into adulthood. The Mental Health Commission of Canada (MHCC) cited the work of Arnett on emerging adults and also used the term “transition-aged youth” to characterize the same age category. The MHCC released a report in 2012 titled “Changing Directions, Changing Lives” where they emphasized the importance of mental health care across the lifespan from youth into adulthood. Despite the variation in start and end points of these developmental periods, there are underlying patterns which suggest youth transition into adulthood over a period of several years. Subsequent theories put forward by Arnett extended the range of emerging adulthood to encompass 18 to 29 years of age and identified an array of mental health implications associated with five major defining traits of emerging adulthood. In this review, Arnett proposed that identity explorations, instability, self-focus, feeling in-between, possibilities and optimism as five potential factors with mental health implications for emerging adults in North America. In Arnett’s review of these five factors, he described emerging adulthood as a period of heightened risk for anxiety and depression associated with instability in relationships and employment, frequent changes of residence, low perceived social support, and the feeling of not actually being an adult yet. Although Arnett’s work was primarily theorizing about the possibility of “emerging adults” as a developmental period between adolescence and adulthood, the use of a term such as “youth” may be more appropriate for general use. For the remainder of this study,
the term “youth” will be used to refer to young people between the ages of 15 and 24 years of age.

1.2 Mental Health of Youth

It is estimated that 1 in 5 Canadians are living with a mental illness and approximately one million are people between the ages of 9 and 19. Additionally, the highest incidence of new cases of mental illnesses tends to be observed in youth between the ages of 12 and 24. Unfortunately, many studies which have characterized mental illness in the population have focused primarily on adults. Youth in general represent a demographic that is currently underserved and understudied, but are characterized by high rates of mental illness and substance use disorders. Among youth, depression and anxiety tend to be some of the most commonly reported mental health conditions. Based on data from the Canadian Community Health Survey (CCHS), one study revealed the lifetime prevalence for depression and anxiety among Canadian youth to be approximately 10% and 12% respectively. This study had a subsample of over 5000 youth from a larger sample pool of nearly 37000 participants, with the CCHS playing a key role in the routine collection of mental health data in Canadians.

Historically, anxiety and depression have presented challenges to researchers due to their similar patterns of prevalence and symptom presentation. Burke et al. (1990) published a book which summarized the literature available at the time on the comorbid nature of mood and anxiety disorders. Firstly, “comorbidity” as defined by Feinstein is the presence of additional (more than one) coexisting ailments within a patient. Burke et al. (1990) cited the work of Freud, who stated that most individuals will present symptoms that are shared across depression and anxiety, and that the two conditions may share a similar etiology. Following this, Clark et al. (1991) proposed development of a tripartite model of mood and anxiety disorders to
address the overlapping nature of depression and anxiety diagnoses. In their tripartite model, the authors state that depression and anxiety have many overlapping characteristics but can be broken down into factors that make them distinct. Anxiety is distinguished from depression by the presence of physiological hyperarousal, or the physical symptoms associated with anxiety such as an elevated heart rate. Depression is distinguished from anxiety by the presence of anhedonia, or the inability to feel pleasure and loss of motivation. The authors state that their proposed tripartite is comprised of specific anxiety (physiological hyperarousal), specific depression (anhedonia), and finally general distress, otherwise known as the area in which depression and anxiety overlap. The authors also proposed that a mixed anxiety-depression diagnosis be used when a discrete diagnosis is not clear. This study is important to the background of the current study as it relates to recently developed approaches for detecting emerging forms of mental illnesses. Trans-diagnostic approaches, which will be discussed in an upcoming section, allow researchers and clinicians to view anxiety and mood disorders more broadly while also appreciate the similarities at the core of the conditions.

1.3 Classification and Diagnosis of Mental Illnesses

1.3.1 The Diagnostic and Statistical Manual for Mental Disorders

The Diagnostic and Statistical Manual for Mental Disorders 5 (DSM-V) is the current tool used by clinicians to provide formal diagnoses of mental, emotional, and behavioural disorders for patients. Across its five iterations, the DSM has seen various changes in the criteria required for a diagnosis as well as the addition and removal of entries due to an evolving understanding of mental illness in the context of society.

The origins of the DSM can be traced back to the World Health Organization’s (WHO) International Classification of Diseases (ICD), an international standard for the diagnosis and
reporting of diseases.\textsuperscript{16} In 1949, the ICD-6 was published by the WHO and was the first version to include the diagnostic and reporting criteria of mental disorders.\textsuperscript{17} This iteration of the ICD was based on the criteria put forward by the Veterans Association following World War II which compiled outpatient data on soldiers who served in the war and any resulting disorders in behaviour or character.\textsuperscript{17} The American Psychiatric Association would later go on to publish the first version of the DSM in 1952, which was heavily influenced by the ICD-6.\textsuperscript{18}

Changes in society’s overall perception of mental illness have impacted which conditions are added or removed from the DSM; however, each subsequent publication of the DSM has resulted in an overall greater number of total entries. For example, in 1973, the DSM II was published which included the removal of homosexuality as a mental illness.\textsuperscript{19} This resulted in a ripple effect which saw the removal of homosexuality from the ICD in 1990.\textsuperscript{19} The parallels in changes to the DSM and changes in societal values are important as they may suggest that societal perceptions of mental illness are malleable. Having diagnostic criteria for mental illnesses that are adaptable and informed by changes in society are also important, which the current study will explore further in the discussion of trans-diagnostic approaches.

Prior to the DSM-V, someone who was experiencing grief due to death could not be diagnosed with major depressive disorder. The removal of this stipulation, what was known as the bereavement exclusion, helped shift society’s understanding of how individuals experience grief and that this experience can overlap with depression.\textsuperscript{20} This change was mostly relevant to the diagnosis of depression, but the underlying issue of overlapping diagnostic criteria extends to various other DSM conditions. Another example that was previously discussed is that anxiety and depression tend to be commonly diagnosed at the same time. This phenomenon will be revisited throughout the current study as the measurement of trans-diagnostic symptoms of
anxiety and depression among youth is an important consideration when screening youth for mental illness.

### 1.3.2 Trans-Diagnostic Symptom Clusters

In an attempt to understand the overlapping nature of symptoms and diagnostic criteria for mental illnesses, some researchers have grouped similar mental illnesses into trans-diagnostic symptom clusters. Buckholtz et al. (2012) proposed that the trans-diagnostic nature of mental illnesses may have biological underpinnings observable in diagnostic imaging of brain regions. Their work suggested that genetic and environmental factors which disrupt certain neural pathways produce manifestations of broad psychopathology as opposed to discrete mental disorders. They proposed that in order to enhance the classification of mental illnesses we must first understand how broad risk-related characteristics of mental illnesses represent a common experience prior to the development of a diagnosis. By proposing a possible biological framework to understand the trans-diagnostic nature of mental illnesses, the work of Buckholtz et al. (2012) informed the potential need for diagnostic criteria to be adapted to accommodate broader trans-diagnostic symptom clusters. By focusing more on broad symptom clusters as opposed to the traditional clinical manifestations of disease, researchers and clinicians may better understand the early phases of a mental illness prior to attaching a name to the illness.

Weersing et al. (2012) focused on studying the “internalizing cluster” in 2012, which was comprised of depression, anxiety, and other somatic complaints. They highlighted the importance of treating the internalizing cluster early on, as the comorbidity of these conditions can greatly harm youth and track with them into adulthood. Their study aimed to create a general set of tools that could be used to treat the internalizing cluster using a trans-diagnostic approach. Their strategy involved going beyond using medication as a simple remedy and including
behavioural therapy in the context of two psychosocial programs. Chu (2012)\textsuperscript{23} echoed these findings in a 2012 report discussing the current status of translating trans-diagnostic approaches to youth and adolescents. Chu explained how the internalizing toolkit could be used to target anxiety-specific or depression-specific symptoms while also being delivered through a therapeutic program that could be tweaked depending on the individual.\textsuperscript{23} The importance of translating mental health interventions to encompass trans-diagnostic approaches is relevant for the current study which aims to examine the sensitivity and specificity of a screening tool in an integrated youth health service. By moving away from compartmentalized approaches for the diagnosis of mental illness it is clear that general interventions can be adapted to individuals and what they are experiencing. This is relevant for screening and delivering integrated care to youth which will be discussed in an upcoming section. This potential approach relies on integrated youth health services which can provide a wide scope of care, and an amended model of staging the progression of mental illnesses. A proposed clinical staging model for mental illness which encompasses the non-specific early stages of mental illness up to the specialized needs of the late stages of mental illness will be discussed in the next section.

1.3.3 Clinical Staging Model of Mental Illness

A recent paradigm shift in psychiatry has focused on better characterizing the progression of mental disorders from youth into adulthood through the creation of a clinical staging model. From oncology, a clinical staging model helps to present the stages of cancer based on the progression of the disease and if it has spread to other parts of the body.\textsuperscript{24} Given that the clinical staging of models of common cancers are well known, clinicians can also determine the appropriate intervention at each stage as well as the chances of survival. McGorry et al. (2007)\textsuperscript{25} have proposed a clinical staging model for mental illnesses which outlines specific stages, the
target population at each stage, and the recommended level of intervention. Where clinical staging models for cancer are made definable by clear changes in the body, the proposed clinical stages for mental illness are slightly ambiguous. Moreover, the first stages of mental illness are defined by symptoms and risk for developing a psychiatric disorder (Stages 0 to 1), whereas the latter stages are defined by the onset and severity of psychiatric episodes as well as the degree of disability (Stages 2 to 4).\textsuperscript{25} A strength of this proposed model would be that the experience of a mental illness is mapped out linearly and suggests some fluidity between the risk of developing a mental illness and the experience of its associated symptoms. With regular screening, which will be discussed later in this section, the proposed clinical staging model for mental illness would aid in early detection and treatment.\textsuperscript{26} If a clinician or researcher identified that a young person was at an early clinical stage for a mental illness, this may help in attaching meaning to their non-specific symptoms and setting the stage for timely and appropriate treatment.

This model has been supported by Scott et al. who stated that the model had potential and endorsed it is a practical tool for psychiatry. In a 2013 study, Scott et al. applied the basic principles of the McGorry staging model to bipolar disorder. The authors commented on some difficulties with applying the model due to the overlapping pattern of symptoms that bipolar disorder can share with depression and other mental illnesses. Following this study, the authors directed the focus of future research towards developing clinical staging models for youth with emerging bipolar disorder. They suggested that the trans-diagnostic nature of a clinical staging model would help contextualize a complex and long-term condition. In the context of the current study, the work of Scott et al. (2013) will be important foundation. Besides proposing ways in which researchers and clinicians can view mental illness broadly, clinical staging models are important for youth and understanding the progression of their illness. In a clinical staging model, trans-diagnostic symptoms will begin to emerge at Stage 0 which is when having tools
that screen for non-specific symptoms will be important. Beyond Stage 0, the use of a clinical staging model to monitor youth with emerging mental illnesses will ensure that appropriate measures can be taken early on and ideally prevent severe disability even in later stages. Clinical staging models must also work in conjunction with two other factors to be successful: integrated youth health services and mental health screening in youth. The upcoming sections will discuss the concept of integrated care and how it aligns with trans-diagnostic approaches to the identification of mental illnesses, and screening for mental illnesses in youth using validated tools and clinical assessments.

1.4 Integrated Youth Health Centres

The World Health Organization (WHO) defines integrated care as a term which contrasts “episodic” or “fragmented” care; instead, integrated care is described as being “seamless”. This report from the WHO also indicated that there is no single concrete definition of integrated care, which would indicate the further need for research on this term. Despite the absence of a formal definition, integrated care usually involves the collaboration of many stakeholders in the healthcare field coming together to provide client-centered care. A definition provided by Leutz states that integrated care seeks to connect the healthcare system with other human services and institutions to improve health outcomes.

McGorry et al. (2007) published an article which identified the need for integrated youth mental health services in Australia. In this article, McGorry discussed that youth mental health services are currently inadequate due to when the transition from youth to adult services occurs. As previously discussed, a common definition of “youth” ranges from 12 to 24 years, and this definition includes the transition from youth to adult mental health services at the age of 18. McGorry cited headspace, an Australian integrated youth health service, as an important
resource in bridging the gap between youth and adult services and providing youth-friendly care. An integrated youth health service is typically a centre which provides inter-professional care and resources to the population it serves. Foundry, which will be discussed later in this study, is an example of an integrated youth health service in British Columbia. Within a centre, one may be able to access primary medical care from a physician or nurse, psychology services and counselling, help from a social worker, and peer-support from fellow clients. Youth from underserved populations may greatly benefit from integrated youth health centres which aid in the early detection and treatment of the greatest causes of mortality among youth such as depression and substance use. Recent studies on headspace have shown the client satisfaction of integrated mental health services and reported favourable results, urging greater research on the effectiveness of integrated mental health care for youth.

1.5 Mental Health Screening in Youth

Mental health screening by researchers or clinicians has been described as a two or three step process, in which individuals are screened for a general likelihood of having a mental health concern, and following screening these individuals are formally assessed by a physician and given a formal diagnosis. The Centre for Addiction and Mental Health (CAMH) provides a summary of the screening process and benefits of screening for mental illnesses in a 2009 research report. In their report, Rush et al. (2009) stated that screening youth for mental illnesses can be beneficial for the identification of co-occurring disorders, early onset of substance use and mental health problems, and to ensure the success of early intervention.
1.5.1 Screening Youth in Schools

Researchers have indicated their support for regular mental health screening of youth in settlings like schools. Ashford (2005) endorsed screening youth for mental illnesses from the position of a schoolboard and stated its importance in supporting the early detection of mental illness, decreasing youth suicidality, and promoting overall mental health. This report was written in response to a screening program that was implemented by Columbia University in 2004, called the Columbia Suicide Screen (CSS). A sample of 1729 adolescents who were suicidal and non-suicidal underwent screening using the CSS and it was determined that odds of suicide increased with the number of emotional symptoms, the number of formal diagnoses, and age. The CSS tool was also validated in a subsequent study by Shaffer et al. (2004), and it was reported that the tool had acceptable sensitivity (75%) and specificity (83%). The implications of this study were that the authors believed the implementation of a secondary assessment would mitigate the effects of a low specificity during primary screening, a finding that will be discussed later as it is relevant to the current study. In addition, a methodological strength of this study is the validation of a screening tool for youth and reporting sensitivity and specificity of the tool as it pertains to youth. Later in this study, the importance of youth-specific screening tools will also be explored.

Dowdy et al. (2010) described the role that screening youth for mental illnesses in schools could play in translating the approach to mental health care from reactive to proactive. The model used in schools was described as being individualistic, and is reliant on high risk youth to be identified and subsequently focus on symptom reduction. Even though this model of care may be well intentioned, a 2003 study by Costello et al. reported that 20% of youth in the United States experience mental health symptoms that would qualify them for a clinical diagnosis of a mental illness. In response to this, Dowdy et al. (2010) commented on the need
for funding better mental health initiatives in the United States by restructuring the current model of mental health services in schools. At the time of the study, it was discussed that schools in the United States needed to switch their focus from providing reactive resources to the most critical cases to proactive universal screening which could aid in the early detection of mental illness. In doing so, they proposed that routine screening of youth in schools would better emphasize the importance of mental health as a public concern and make a health impact at the population level. Although mental health screening in schools is not contextually relevant for the current study, it is still an important background area to explore. This is because in Canada youth under the age of 16 or 18, depending on the jurisdiction, and are legally required to be in school. Schools as research settings are important to consider when wanting to study youth under the age of 18; however, for youth between the ages of 18 and 24 clinical settings may provide more information on service usage and results from clinical assessments. The next section will discuss mental health screening in clinical settings and its important in the context of the current study.

1.5.2 Screening Youth in Clinical Settings

In the past, it was reported that the treatment of mental illnesses took place in primary care sectors as opposed to mental health sectors of the healthcare system. Shapiro et al. (1984) summarized the service utilization of adults over the age of 18 in three catchment areas, and described the differences in seeking primary care or mental health specialists for treatment. Their sample consisted of approximately 9000 individuals, and the 6-7% of the sample who sought after mental health services in the previous year. They reported that of these individuals with DSM-III diagnoses or severe cognitive impairment, only 24% to 38% across the sites sought out specialist mental health care. Alternatively, for individuals who seek mental health care through the primary care sector, it was common for mental illnesses to be underdiagnosed by
primary care physicians. Schulberg et al. (1988) stated that upwards of 25% of youth primary care patients present with mental health concerns, and cites theories from other studies that may explain why people with mental health concerns may go underdiagnosed in primary care settings. Schulberg et al. (1988) highlighted that rates of depression diagnoses vary widely and cited Mann et al. (1981) who suggested that the reason why rates of diagnosing depression vary was due to the misclassification of depression as anxiety. In response to this, Schulberg et al. (1988) recommended that future work focus on better distinguishing anxiety and depression; however, given the context of the current study, this finding further underscores the importance of trans-diagnostic approaches and the development of clinical staging models. It is unclear whether the research summarized by Schulberg et al. was conducted on youth, but another area worth exploring is the underreporting of mental health concerns in primary care settings due to perceived stigma.

In a 2002 editorial, Sartorius called upon psychiatrists to consider the barriers that stigma imposes on those with mental illness to improve their health. In this article, the author discusses the frivolous use of labelling mental illnesses and the treatment of patients by mental health professionals as key sources of stigma. Regarding labels, the author comments on their helpfulness in attaching meaning to a collection of symptoms someone may have but also explained how they can become less useful when their meanings are misinterpreted by the public. Sartorius (2002) also explained how patterns of behaviour displayed by psychiatrists who request greater compensation and time off for working with “dangerous” patients only added to the stigmatization of people with mental illnesses. Sartorius recommended that psychiatrists think carefully about their comments and actions, and how this can contribute to the societal stigma of mental illness.
At this point, there are several patterns that have emerged from the background literature that help frame the current study and the next section which introduce the screening tool of interest. Historically the access to mental health care has been decentralized, there have been patterns of misdiagnosis or unrecognition of mental illnesses, and clinical environments that have not been inviting to those seeking care. Further research is required to understand the specific impact these issues have had on youth. The current study is attempting to focus on potential solutions to the previous issues described. First, integrated youth health services would help in providing centralized access to mental health resources. Second, the creation of youth-friendly environments that bridge the gap between adult services that aim to decrease stigma. Third, using trans-diagnostic approaches and the clinical staging of mental illness to make sense of non-specific illnesses or the overlapping nature of depression and anxiety. In the next section, the GAIN family of instruments will be introduced and how they can potentially address some of the issues identified with screening youth in clinical settings.

1.6 The GAIN Family of Measures

1.6.1 Clinical Interviews, Assessments, and Screening Tools

The GAIN family is comprised of several instruments, each of which may be used in a different population, administered by a different individual, and have a different purpose or outcome. In order to thoroughly explain the nuances of the GAIN it is important to provide definitions for clinical assessments and screening tools. Simply put, a clinical assessment is usually conducted by a clinician, such as a physician or nurse, and typically results in a formal diagnosis of a disease or condition. For example, by using the diagnostic criteria provided by the DSM-V, a physician can perform a clinical assessment on a patient and arrive at a formal diagnosis of a mental illness. A clinical interview is the process in which a clinician asks a
patient the questions outlined in a clinical assessment to arrive at a diagnosis. A screening tool can be administered by a clinician, researcher, or self-administered depending on the tool and may be completed by clients prior to or during a clinical assessment. As previously discussed, the purpose of using a screening tool is to categorize individuals into some level of risk but not provide a formal diagnosis. The GAIN family contains several tools which function as full clinical assessments as well as others that are used for brief screening.

1.6.2 The Global Appraisal of Individual Needs

The Global Appraisal of Individual Needs-Initial (GAIN-I) is a clinician–reported clinical assessment tool that has been used as a semi-structured assessment to provide a formal psychiatric diagnosis.\textsuperscript{44} The GAIN-I will simply be referred to as the GAIN for the purpose of this study due to it being the most comprehensive of the GAIN family. The term “initial” is derived from the use of the GAIN as an initial assessment to determine resource and program allocations for the interviewee.\textsuperscript{45} The GAIN is a tool that has been used as a semi-structured assessment, a hybrid of a clinical interview and a highly structured assessment. In its long form, the GAIN serves as a diagnostic tool which asks clients a series of questions if they are experiencing problems with substance use or mental and emotional health, involved in the criminal justice system, and questions about physical health.\textsuperscript{44} Due to the significant breadth of items, the GAIN assessment can take two hours to administer.\textsuperscript{44} Over time, the GAIN family of instruments has expanded to include various tools designed for different usages and settings.\textsuperscript{45}

The GAIN Monitoring 90 Days (GAIN-M90) is a GAIN tool that was developed for the purpose of monitoring an individual following initial assessment. In clinical and research settings, the GAIN-I and GAIN-M90 may be used alongside each other, where the M90 is administered quarterly for the purpose of collecting follow-up data after the administration of the
Overall, it is different from the GAIN-I due to the fact that it is not a stand-alone assessment and must be completed subsequently after an initial assessment.

The GAIN-Q3 is a brief assessment tool which collects information on substance use, internalizing problems and mental health concerns, crime and violence problems, as well as other physical, social, and environmental health information and can be completed in under an hour. The GAIN-Q3 is similar to the GSS, as they are both considerably shorter than the GAIN-I in length and time of administration. The difference between the GAIN-Q3 and the GSS lies in their intended usage, in that the GAIN-Q3 is described as a “brief assessment” and the GSS is a screening tool for the general population. Ultimately, the GAIN-Q3 can be thought of as an intermediate tool between a full clinical assessment (GAIN-I) and a screening tool (GSS). The GSS will be discussed at length in the following section and will form the focus of the main analytical sections of the current study.

1.6.3 The GAIN Short Screener (GSS)

One of the most widely used variants of the long form of the GAIN, is the GAIN Short Screener (GSS). The GSS is a 23-item screening tool which measures significant problems across the domains of internalizing disorders, externalizing disorders, substance use, and crime and violence problems. Internalizing disorders can include anxiety and depression, and externalizing disorders can include Attention Deficit Hyperactivity Disorder (ADHD) and Conduct Disorder (CD). Together, internalizing and externalizing disorders have been described as both major domains of psychopathology. In the GAIN, problems with substance use can include problematic use of alcohol, marijuana, and other illicit substances. Finally, crime and violence problems are concerned with the participants’ involvement in illegal and criminal activity, and their involvement with the criminal justice system. The measurement properties of
the tool will be discussed at length in the following chapter of this study. The GSS garnered attention for being a rapid tool that was able to capture both major domains of psychopathology (internalizing and externalizing disorders) as well as substance abuse and crime and violence problems. The work of Dennis et al. (2006)\textsuperscript{47} was important in establishing the validity of the GSS, and provided the foundation for the current study to continue to explore the tool.

In the context of an integrated youth health centre, there are various advantages associated with the use of a tool such as the GSS. It provides a rapid screen that is easy to administer at an intake assessment or primary care visit. If a client is seeing a physician during their visit, there will be an immediate flag on their profile if they screen positively. Second, the GSS covers a breadth of behavioural disorders. Essentially, it can replace the use of several separate screening tools and in so doing improves the efficiency of the clinical environment. Although the GSS cannot provide a formal diagnosis, if an individual screens positive on the GSS they would proceed to a clinical assessment which frequently includes administration of the full version of the GAIN.\textsuperscript{47} This process of screening individuals with the GSS and then administering a full clinical assessment of the GAIN was the primary methodology used to validate the initial version of the GSS. This initial validation of the GSS helped to establish important facets of the screening tool’s validity including sensitivity and specificity as well aid in determining clinical cut points.

1.7 Psychometric Properties of Clinical Assessments and Screening Tools

1.7.1 Validity

In quantitative research, there are many different forms of validity, but in general the term is used to describe the ability of tool to accurately the measure what it is intending to measure.\textsuperscript{48} Internal validity, generally refers to the accuracy of a test, tool, or instrument in being
able to answer a research question. Measures of screening test validity include sensitivity and specificity which will be discussed at length in this chapter. External validity, otherwise known as generalizability, refers to the extent at which the findings in a research study can be applied to the general population. Factors that influence external validity include the sample size, the study design, and the sampling method that was used in a study. Measures of validity for screening tools are particularly important because they inform researchers and clinicians if the tool is fulfilling its purpose. They are also important in assuring patients and clients about the accuracy of test results and how these may impact their health. In addition, it is important to have multiple and varied validity studies for a tool as its accuracy may fluctuate in different environments and thus affect test results. The GSS has been validated in the past and the current study aims to further contribute to the literature by examining the sensitivity and specificity of the tool.

1.7.2 Reliability

Measures of reliability are distinct from validity but are still important in the development of a strong screening tool. Where validity measures the ability of a test to do something accurately, reliability is important to ensure that a test can do something consistently. Ideally, a strong screening tool will measure something accurately and consistently. Most measures of reliability provide an estimate of the extent to which two raters or two instruments agree, and these measures will also be discussed later in this study. In combination with validity, this study also aims to provide some measures of reliability of the GSS in the context of an integrated youth health service.
1.8 Objectives and Study Hypotheses

The first objective of this study was to review the GAIN and GSS literature. This involved establishing a set of search databases and search criteria and performing a screening process to filter publications by title, abstract, and then finally the complete article. In Chapter 1.9, the results from this detailed literature search are presented based on how they implemented the GAIN or GSS.

The second objective of this study was to contribute to literature on the validity of the GSS as a screening tool by examining its sensitivity and specificity. The development of Receiver Operator Characteristic (ROC) curves, and the determination of optimal cut points for the tool was also completed. Regarding this objective, it is hypothesized that the results will be consistent with previous studies indicating the GSS has acceptable sensitivity and specificity greater than 80%.

The third and final objective of this study was to contribute to literature on the validity of the GSS internalizing disorder screener (GSS-I) by examining its sensitivity and specificity using a similar methodology. This appears to be the first study to validate the GSS-I independent of the rest of the tool, but it is hypothesized that a similar pattern of findings to the GSS will be observed for the GSS-I. The results of both the analytical objectives of this study are presented in Chapter 3.

1.9 Literature Review of the GAIN and GSS

1.9.1 Literature Search Strategy

This literature review was conducted using a structured search strategy which involved the use of multiple databases, with supplemental searches completed using Google Scholar. The purpose of this literature review was to identify and summarize studies which have used the
GAIN or GSS as measurement tools, validated the GAIN or GSS, or used the GAIN or GSS as gold standards to validate another measurement tool.

1.9.2 Databases

The following databases were accessed to perform this literature search: EMBASE, EBSCOhost, Web of Science, PsycINFO, PubMed, Ovid Medline, CINAHL, and ERIC. The search was completed entirely on February 1, 2019 and all articles returned from the search were saved.

1.9.3 Article Screening

On every database the term “Global Appraisal of Individual Needs” was searched and a total of 233 results were exported to Mendeley. On Mendeley, duplicates were removed which brought the total number of articles to 168. Following this, titles and abstracts were screened together and studies which did not include the GAIN or GSS in some capacity were excluded. Other criteria such as age, sex, and country were not used as exclusion criteria because the purpose of this search was to cast a wide net and summarize as many available articles as possible. The screening of titles and abstracts was combined to prevent the removal of studies which did not include their measurements in the title of the publication. Following this first stage of screening the total number of articles was reduced to 96. The second stage of screening involved pulling full articles to determine their relevancy which brought the final article count to 93. Full articles were excluded if they did not include the GAIN or GSS in some capacity.
1.9.4 Article Categorizing

Studies which utilized the GAIN and the GSS were subdivided into four different types of studies: 1) studies which used the GAIN or GSS as assessment or screening tools, 2) studies which examined the validity of the GAIN or GSS, 3) studies which used the GAIN or GSS as gold standards, and 4) studies which fell outside of the three aforementioned categories.

1.10 Literature Search Results

The results of the detailed literature search that was conducted for this study are presented in the following section. Several studies were chosen to be highlighted in this section to discuss their methodology and relevance to the current study. In Figure 1 a diagram is provided to describe the process that was taken to search online databases and arrive at the articles reviewed for this study. Additionally, all studies from this literature search which were available online were compiled in Appendix A. The appendix lists each study by their title, authors, date of publication, and includes a brief summary.
Figure 1 - Diagram outlining the process that was taken to arrive at the results of the literature search including the stages which screened articles and then categorized them based on how they utilized the GAIN or GSS.
1.10.1 Studies Which Used the GAIN and GSS as Assessment or Screening Tools

The work of Dennis et al. (1998) on the development of the first version of the GAIN dates back over 20 years, but it appears that the first study to use the GAIN as a tool for measurement was published by Tims et al. in 2002. The GAIN was used for the purpose of diagnosing Substance Use Disorder (SUD), Conduct Disorder (CD), and Attention Deficit Hyperactivity Disorder (ADHD) in adolescents presenting to outpatient cannabis abuse treatment centres. The authors also measured other co-occurring disorders that are found in dimensions of the GAIN and the diagnostic criteria found in the DSM-IV. This initial study was followed promptly by Dennis et al. (2002) in the same year, where the GAIN was also used for the purpose of measuring outcomes at baseline and at follow-up assessments. In addition to being used as an outcome measure, Dennis et al. (2002) also utilized an innovative methodology which would evaluate the financial implications of the elements of the GAIN to provide each participant a unique benefit and cost ratio which would then be used to perform an economic analysis of resource utilization of the research site.

Between 2002 and 2018, it appears that the GAIN family of instruments were used in 60 studies as clinical assessment tools or screening tools (Appendix A). Although the majority of these studies were unique, there were many common themes that were consistent across the settings of the study, the methodologies used, and the outcomes. First, studies which utilized the GAIN or GSS were typically based in an outpatient treatment facility, usually for the treatment of substance use related issues. Second, the GAIN or GSS were usually employed as the primary assessment or screening tool or supplemented with other tools to collect data pertaining to outcome measures. If the GAIN was used, it was for the purpose of performing a clinical assessment and determining if participants met the diagnostic criteria for a mental illness. In some instances, the entire version of the GAIN was not administered, but various subscales of
the full GAIN may have been specifically chosen which were most relevant for the study question and outcomes. This reduced the overall time of administration for the GAIN and allowed for self-administration if the researchers were only screening using specific subscales of the GAIN. Finally, if the GAIN was being used to produce longitudinal data where it was typically administered at baseline during patient intake or at the start of the research study and repeated in subsequent rounds of data collection of three- or six-month intervals. If the GSS was used it was for the general purpose of screening across the four major domains of mental and behavioural problems measured by the GSS to categorize participants into a low, moderate, or high-risk category. The following studies were summarized for their use of the GAIN or GSS in innovative ways that demonstrated their wide use as screening or assessment tools.

Lennox et al. (2006) combined the psychometric measures of GAIN (Substance Frequency Scale) with biometric measures (urine and blood) to build structural equation models to understand what factors underlie substance use.\(^{52}\) The main findings from this study highlighted that the GAIN’s Substance Frequency Scale (SFS) actually performed better in all the problem areas measured relative to the biometric measures. The author’s commented on the shortcomings of biometric measures being that they are not sensitive enough to discriminate levels of severity of substance abuse relative to a psychometric screening tool. Although future studies would see the development of screening tools designed for screening co-occurring disorders, a study by Chan et al. (2008) was one of the earliest uses of the GAIN to explain comorbidity of significant internalizing and externalizing problems among adolescents presenting to substance use treatment facilities.\(^{53}\) One methodological strength of this study was the use of a large sample of adolescents (n=4930) and adults (n=1956) who were then divided into five different age categories to study the relationship between comorbidity and age. The Chan et al. (2008) study found that youth had the greatest number of co-occurring mental health
conditions, and youth with co-occurring conditions were also more likely to have problems with substance use as well. Co-occurring disorders can be thought of as comorbid disorders, as previously defined, comorbidity describes instances in which an individual is diagnosed with multiple psychiatric conditions. The findings from this study are important as they highlight the health service requirements of youth populations presenting to substance use treatment and the usefulness of the GAIN as an assessment tool used in this context.

Womack et al. (2004) used the GAIN alongside the DSM-IV, where a two-step screening process was used to flag participants with depression. In this study, participants were initially screened with the GAIN for depression. If the participant screened positively on the GAIN for depression then they were provided with a follow-up assessment using a DSM-IV diagnostic criteria for depression. If the participant screened positively on the DSM diagnostic criteria tested positive, they were invited to participate in the intervention for the study. Although this method of utilizing the GAIN is not common, the researchers may have felt that the combination of the GAIN and the DSM criteria would provide the most accurate assessment of depression.

Due to the length of the complete GAIN, it has typically been administered in the form of a clinical interview by a clinician or researcher. Titus et al. (2012) studied the impact of training to certify individuals to administer the GAIN. They determined that the fidelity of administering tool generally increased with training and it was recommended that an individual administer the tool 15 times prior to being certified. This appears to be the only study which involved using the GAIN as an assessment tool while also certifying individuals to conduct the assessments as part of the study. Although the GAIN is useful as an assessment tool, following this study it became clear that in order for the assessment to be the most effective the involvement of a certified administrator is important. In addition, the length it takes to administer
the GAIN may not always be the most convenient. The GSS can effectively aid in addressing these potential limitations through the implementation of a two-step screening procedure. Given that the GSS can be administered without prior certification, it can be used first to screen for their risk of mental or behavioural problems. Subsequently, patients can receive a complete clinical assessment with the GAIN should they screen positively on the GSS.

Studies which used the GAIN as a screening tool may have opted for subscales within the full version of the tool as opposed to the full assessment; however, a study by Tarantino et al. (2013) was one of the first to use only one item from the GSS to screen participants.56 This study used one item from the GSS to assess suicidal ideation in participants in a study which examined the mediating effects on familial relationships associated with suicidal ideation among substances users. The authors’ choice for this was rooted in previous research suggesting that single items which assess suicide ideation and are dichotomous (two options) are valid assessments for this type of focused screening.56

The support of the GAIN and GSS as mental health screening tools in Canada is significant, as the tool has been approved for standardized use across Ontario (via CAMH)33 and in Alberta (via Alberta Health Services)57. A national, multi-site, randomized trial published its findings on housing first programs across Canada in 2015, in which the Substance Problem Scale (SPS) of the GSS was used to determine problematic substance use among participants.58 Although the GAIN and GSS are not standardized for use across all of Canada, this study by Aubry et al. (2015)58 is one of few to support the use of the GSS in a country-wide study. This study was consistent with previous studies where only the substance use scale of the GSS was used as a screening tool. This may be due to the presence of literature which has endorsed the use of the SPS for use independent of the rest of the tool. As previously mentioned, the usage of the GAIN and the GSS in research settings in the past has usually been administering the tool at
intake or as participants present to substance use treatment which was once again observed in this study.

An emerging use of the GAIN, demonstrated initially by Aubry et al. (2015), is in research focusing on precariously housed youth. In 2017, separate studies by Csiernik et al. (2017) and Smith et al. (2017) used the GSS in investigations of precariously housed youth and housing instability in Canadian contexts respectively. In the Csiernik et al. (2017) study, the SPS scale of the GSS was used to determine if participants endorsed problematic substance use. In the Smith study, a more substantial commentary on the use of the GSS in the context of measuring housing instability was provided. The authors indicated that their study would inform future research if they were able to demonstrate the scores on the GSS were elevated among precariously housed youth. The researchers found that overall precariously housed youth presented to clinical settings and endorsed greater mental health issues relative to non-precariously housed youth. Additionally, the researchers noted that the majority of youth presented to non-clinical sectors (such as educational and justice sectors), and the authors recommended that routine screening should be occurring across a range of service sectors that support youth. The significance of these recommendations is that the authors used the GSS in an emerging area of research and suggested that regular screening should be more widely used in this field. One of the most recent studies in this regard was by Barnes et al. (2018) which reported consistent findings with previous studies, youth experiencing family homelessness would be at a greater risk for various mental health concerns.

After reviewing the GAIN and the GSS literature there is a clear history of usage in outpatient treatment facilities for substance use, and this finding is relevant for the current study which was conducted within an integrated youth health service. By using previous literature as a
reference point, it would appear that the use of the GSS in the current to screen youth in an integrated health service was appropriate.

1.10.2 Studies Which Used the GAIN and GSS as Gold Standards (Reference Tools)

When selecting a gold standard, especially if there is more than one widely used tool within an area of research, it is important to evaluate the psychometric properties of the tool. Researchers must consider the validity and reliability of the gold standard. Generally, if an instrument is being used as a reference tool then it must have demonstrated an excellent level of accuracy and consistency in its measurement. In the next section, studies which have used the GAIN and GSS as gold standards will be discussed. Later on, a discussion on the use of the terms “gold standard” and “reference tool” will be provided. For the purpose of this study the term “reference tool” will be used to refer to a tool that filled in the role of a gold standard in the measures.

It appears that the first study to use the GAIN as a gold standard was by Titus et al. (2003). The purpose of this study was to validate the General Victimization Index (GVI) which was a 15-item with the GAIN-Initial (GAIN-I) and the GAIN-Monitoring 90 Days (GAIN-M90). Findings from this study indicated that the GVI is an internally reliable scale with a Cronbach’s alpha statistic of 0.88. Although this was a very focused study since it only utilized the GAIN as a reference tool, the use of the GAIN family of measures as reference tools remains important.

The Criminal Justice Drug Abuse Treatment Studies (CJ-DATS) were published by Sacks et al. (2007) and involved the validation of two new brief screening tools intended for use in criminal justice settings. In this study, the GSS and two other screening tools were used to validate the Co-Occurring Disorders Screening Instrument for Mental Disorders (CODSI-MD)
and the CODSI for Severe Mental Disorders (CODSI-SMD). These tools were developed and validated for the purpose of screening for serious co-occurring mental disorders and substance use disorders in criminal justice settings. The purpose of the CODSI instruments was to screen for mental health concerns and substance use in a very specific population. It is worth noting that even though the GSS can screen general populations, using it as a screening tool in a criminal justice setting would still be appropriate. When the GSS was initially developed and validated by Dennis et al. (2006) there were participants who were involved in the criminal justice system part of their study sample. This may form part of the reason why the developers of the CODSI instruments decided to use the GSS as a gold standard for validation. The method in which the two versions of the CODSI were validated involved examining the sensitivity, specificity and overall accuracy of the tools, which was a ratio defined by the authors as the number of participants correctly identified or correctly ruled out of having a mental disorder by the tool. The study also generated cut-off scores for the two CODSI instruments using ROC curves. Following the analysis, it was revealed that the CODSI instruments performed reasonably well and comparably to the reference measures used but it was noted that the GSS was one of two best performing instruments in this study. The CJ-DATS studies are important in the history of the GAIN and GSS as they were also utilized in an early study which involved validating the GSS, shortly after the first major validation study of the GSS completed by Dennis et al. (2006).

Following this, the Co-Occurring Joint Action Council (COJAC) developed their own screening tool known as the COJAC Screening Tool (CST) to be used for screening co-occurring disorders in clinical settings. Jessup et al. (2011) validated the CST by using the GSS as a reference tool in their study, similar to the validation of the CODSI instruments. The authors indicated that due to the fact that the GSS has items which are “time-sensitive”, six items on the
CST had to be modified to ensure the tools were more comparable. This was accomplished by asking participants who completed the CST if they experienced the symptoms described in six particular items in the previous month. In a convenience sample of youth (n=1951) the GSS was reported as having greater internal consistency (Cronbach’s alpha = 0.90) than the CST (Cronbach’s alpha = 0.44) for the measurement of co-occurring mental health disorders and substance use problems. In discussing the limitations of the study, the authors indicated that due to some items of the CST being poorly correlated with those on the GSS they may have been altered or removed in subsequent iterations of the tool. This study highlighted important differences in how the GSS measures symptoms, and these nuances will be discussed in an upcoming section.

Veldheizen et al. (2014)65 validated a screening tool which was composed of just one item which asked “Do you think you have mental health problems?” The Single Screening Question (SSQ) was designed to screen for mental disorders in substance use treatment facilities and was intended to be an extremely brief screening tool that would be accurate enough to screen populations presenting to treatment.65 The SSQ was validated with the GSS and other brief screening tools which consisted of five or six items, and participants were also interviewed to determine if they met the criteria for a mental illness. The study examined sensitivity, specificity, and generated ROC curves to provide evidence on the validity of the SSQ. The results of the study demonstrated that the tool was as accurate as the other five and six-item screening tools but was out-performed by the GSS. The main implications of this study are that a one item screening tool was able to perform seemingly as well as other validated tools which are longer. Researchers and clinicians have to consider the merits of such a short screening tool, in that it could significantly improve the efficiency of mental health screening. The main drawback of such a brief screening tool is that it may not be able to provide as much useful information as a
screening tool like the GSS which can inform a clinician or a researcher what specific problems an individual has been experiencing before subsequent stages of screening.

The selection of a gold standard is reliant on that tool having adequate psychometric properties to be able to accurately and reliably identify a positive or negative case. For a researcher or clinician to select the GAIN or GSS as gold standards, they must have looked at existing literature which provided evidence on the validity of the tools. The next section will explore the existing GAIN and GSS literature to provide information on the psychometric properties of the tools.

1.10.3 Studies Examining the Validity of the GAIN and GSS

1.10.3.1 GAIN Validity and Reliability

The GAIN administration manual provides a section summarizing the validation studies that have been published for the tool since its conception; however, the majority of studies which “validated” the GAIN only measured internal consistency and inter-rater agreement, which are measures of reliability. Dennis et al. (2002) validated the GAIN as part of the Cannabis Youth Treatment (CYT) experiment, a study which discussed in a previous section. The main objective of the CYT experiment revolved around providing evidence-based guidelines on five different interventions for cannabis treatment in youth including: Cognitive Behavioural Therapy (CBT), Motivational Enhancement Therapy (MET), Family Support Network (FSN), Multidimensional Family Therapy (MDFT), and Adolescent Community Reinforcement Approach (ACRA). The GAIN’s utilization in this study was two-fold, as it was used in assessing outcome measures and it was also validated through the analysis. The GAIN was the primary measurement instrument used in this study, being administered to adolescents at intake and in three-month follow-up intervals subsequently after intake. The GAIN was supplemented
with other measures and compared to parent-reported measures and biometric measurements. The GAIN was reported to have inter-rater agreement that was consistent with other measures, with Cohen kappa values ranging from 0.70 to 0.90. One of the goals following this study was to monetize measures of the GAIN associated with substance use to assign economic values to each participant, allowing for an analysis of the cost and benefits per adolescent at each site. Although there is little information available on the actual validity of the GAIN, it can be assumed that because the GAIN is a clinical assessment based on diagnostic criteria that it will be valid.

1.10.3.2 GSS Validity

Dennis et al. (2006) were among the first group of researchers to validate the GSS with the GAIN used as a gold standard. The results provided primary evidence to support the use of the GSS to screen for internalizing, externalizing, substance use, and crime and violence problems among adolescents and adults. Immediately, it was clear to researchers that the GSS posed benefits as a screening tool since it was a shorter tool than the full GAIN that could be self-administered and completed in under five minutes. The data used in is initial examination of the validity of the GSS came from 77 different studies and over 30 different sites, with adolescents between 10 and 17 (n=6177) and adults between 18 and 69 (n=1805). To this date, of all the studies to validate the GSS this study by Dennis et al. (2006) appears to have the largest and most diverse sample. In this study the 20 items of the original version of the GSS were used to predict the responses of the 123 items on the full version of the GAIN. Dennis et al. (2006) determined the sensitivity, specificity, and established cut points for the GSS through ROC curves and AUC values. Overall, the findings suggested that the GSS had excellent sensitivity (90%) and specificity (92%) when validated with the GAIN and may have merit as a screening tool to be used in clinical populations of youth or adults. Comments made by the authors in this
study helped establish the role of the GSS as a screening tool, in that the authors acknowledged that the GSS should not be used to replace a full clinical assessment. Instead, it should be used to flag individuals who are at risk, then follow-up with a clinical assessment to confirm a potential diagnosis and determine what resources are necessary for that individual. Dennis et al. (2006) recommended that researchers and clinicians approach mental health screening in two stages: a brief initial screen to identify suspected cases, and a complete follow-up assessment to determine the severity of cases and address complexity and morbidity of cases. It was proposed that the GSS could be the screening tool to accomplish these goals due to its brevity and connection to the full version of the GAIN, which could be administered after the GSS to ensure continuity of the screening program.

The second study to validate the GSS was by McDonell et al. in 2009 and utilized a similar methodology as the current study. This study had a sample of 95 adolescents and examined the psychometric properties of the tool, such as sensitivity and specificity, and determined optimal cut points through the use of ROC curves. Following the analysis, the authors reported the sensitivity and specificity of the GSS as 88% and 89% respectively. On the total disorder screener, McDonell et al. (2009) recommended the use of a cut point of 3 to optimize sensitivity and specificity. In their conclusions, the authors were hesitant to recommend the use of all GSS subscales with the exception of the substance use problems scale. This was due to the lack of adequate validity in the internalizing, externalizing, and crime and violence domains. The authors recommended that the GSS be used to screen in lower-risk populations, such as the general population and primary care settings.

Following the work of McDonell et al. (2009), a study by Smith et al. (2017) validated the substance use problem scale of the GSS independent of the rest of the tool. The sensitivity and specificity of the GSS substance use problem scale were reported to be 83% and 95%
respectively, with a recommended cut point of 2. The gold standard used in this study was the DSM-IV criteria for a substance use disorder, which can be thought of as a clinical assessment resulting in a diagnosis. The authors recommended that future studies validate the GSS with other screening tools to determine its performance relative to other instruments. The current study will attempt to address this gap in the literature by examining the sensitivity and specificity of the GSS-I with other screening tools which measure depression and anxiety.

1.10.3.3 GSS Reliability

Along with their initial validation of the GSS, Dennis et al. (2006) also reported that the total disorder scale of the GSS had good internal consistency (alpha = 0.96), and that it was strongly correlated with the GAIN (Pearson r = 0.94). As previously discussed, the large sample size of this study, the diversity in age range, and the sampling across multiple clinical sites made it stronger in terms of methodology. The work of Dennis et al. (2006) will be discussed again as it formed basis of the analytical plan that would be used to examine the sensitivity and specificity of the GSS and GSS-I in the current study.

McDonell et al. (2009) also analyzed the reliability of the GSS by reporting the internal consistency of its sub-scales by using Cronbach’s alpha. The authors found that all the subscales were highly correlated with the total disorder screener of the GSS, with Pearson’s r ranging from 0.67 to 0.82. Values for Cronbach’s alpha for each subscale were as follows: internalizing scale (alpha = 0.66), externalizing scale (alpha = 0.55), and substance problems scale (alpha = 0.89). The authors did not provide any measures of internal consistency for the crime and violence scale of the GSS. Although this study provided a thorough analysis of the validity and reliability of the tool and its subscales, the small sample size of 95 adolescents may impact the ability to draw definitive conclusions.
For an integrated youth health service to confidently use the GSS to screen its clients, it is important that the tool demonstrate that it is valid and reliable. Although the amount of research on the GSS in this domain is modest, the current study aims to address these gaps as outlined in the objectives section. By conducting a thorough review of the literature on the GAIN and GSS, it is clear that the tool has been validated for youth and adults in clinical settings but continued evidence to support the validity and reliability of the tool is required. By drawing on a sample of youth from an integrated youth health service the current study will be able to provide transferable findings and implications to the use of the GSS in this setting.

1.10.3.4 Other Studies Which Utilized or Studied the GAIN

The following studies have used the GAIN or GSS in a method other than the three that were previously stated. The purpose of summarizing this literature is to provide insight on the various ways the GAIN and GSS have been used outside of a quantitative research lens. One of the earliest entries in this domain was a study by Friedmann et al. (2008) which outlined the implementation process for the GAIN on a state-wide level in Missouri. In this study the authors outline six steps which formed the basis of the implementation process: 1) Exploration/Adoption, 2) Program Installation, 3) Initial Implementation, 4) Full Operation, 5) Innovation, and 6) Sustainability. In order for the implementation of a state-wide program to be successful, the author’s emphasized the importance of having the full support of researchers, staff, and clients as well as state support through funding of sites, personnel, and resources. Ultimately, this appears to be one of the first studies to study the GAIN in a qualitative implementation oriented setting and has implications that can be extended and reproduced by other states or countries looking to implement the GAIN in a standardized form.
The current study focused on the use of the GSS in a Canadian context, but it appears that there has also been international work completed to translate the GAIN and GSS to other languages. Claro et al. (2012)\textsuperscript{70} translated and culturally adapted the GAIN to Brazilian-Portuguese and determined the content validity of the new version. Content validity is a type of validity which measures the extent to which something accurately measures of all facets of a construct. Due to the ability for items to get misconstrued following translation, it was important that the content validity of the Brazilian-Portuguese version be adequate. They determined that the Brazilian-Portuguese version had a content validity of 0.91, which was considered acceptable by the authors, but there was a demonstrated need for continued validation of this new version of the tool. A follow-up study by Claro et al. (2016)\textsuperscript{71} aimed to study the internal consistency of the newly translated version of the GAIN as well as Brazilian-Portuguese version of the GSS. The study drew on a sample (n=168) of individuals in outpatient treatment facilities or community settings, where the average age of participants was 36.8 years. For both translated versions of the GAIN and the GSS the value for Cronbach’s alpha was reported to be equal or greater than 0.70.\textsuperscript{71} Although the cultural adaptation of the GAIN and GSS is not completely relevant in a Canadian context, the support of the GAIN family in other nations is promising and the spread of the tool to other youth health services could be beneficial.

The results of the literature search that fell outside the major scope of this study, provide some important findings that may be relevant to the use of the GSS in a broader population lens. The implementation process of a screening tool is rigorous but may be important for health service providers to consider if they wanted to use the GSS or another screening tool in a standardized format. The recommendations of this study which will be discussed in the final chapter will help provide some guidelines to the province of British Columbia for using the GSS and the role that Foundry could play in that process.
Chapter 2: Research Hypotheses and Methods

2.1 Research Objectives for the Analytical Studies

Based on the research objectives and hypotheses outlined in Chapter 1.8 and informed by the results of the literature search summarized in Chapter 1.10, the following chapter presents the methodology and results of the two analytical objectives of the current study.

2.2 Study Setting and Recruitment

Located in British Columbia, the Vancouver-Granville Foundry, is one of eight centres that currently make up the Foundry network and was the primary setting for this study. Foundry serves as an important setting for this study due to the current need for research on validity of screening tools, in particular the GSS, in an integrated youth health service. The Vancouver-Granville Foundry Centre provides a wide range of health and social services to youth all housed in one building, including primary medical care, integrated case management with a social worker, peer support, and counselling services. Youth were recruited to participate via flyers placed in high traffic areas of the clinic. A member of the research team set up a table in the clinic to answer questions and enrol participants.

2.2.1 Foundry

In 2007 the Inner City Youth program (ICY) was launched by a team of psychiatrists at St. Paul’s Hospital who recognized that youth required better mental health, substance use, and primary care services.\textsuperscript{72} Five years later, donations helped the ICY program move beyond simply providing health services and expanding its scope to help connect youth with their community and peers. Opened in 2015, The Granville Youth Health Centre (GYHC) was the first integrated youth health service in Canada and it housed the ICY program and additional services in the
same building. In order to further these efforts beyond the GYHC, the BC Integrated Youth Services Initiative was established, which would later become known as Foundry. Foundry was inspired by the principle that early intervention of health problems youth may encounter can prevent them from becoming more severe and harming their relationships with peers, families, and education. Today, Foundry has a total of eight sites with three locations across the lower mainland, two locations on Vancouver Island, two location in the interior of BC, and one location in northern BC.

Vancouver-Granville Foundry (VGF) is an integrated youth health service for youth located in downtown Vancouver, British Columbia. VGF provides an array of supports including but not limited to, primary care, mental health and substance use services, psychosocial rehabilitation, peer support, and integrated case management with a case manager for youth between 12 and 24. Foundry’s core principles revolve around the provision of safe and non-judgmental care, and is modelled around prevention and the early detection of health problems before they become problematic. In addition to having a holistic view of healthcare, Foundry’s integrated structure allows clients to access all of their required supports in one space. The centre tailors services to its clients by using a stepped-care model in which individuals with a greater level of need will receive a greater amount of services. Foundry’s services require feedback in order to ensure the highest quality of care. To accomplish this, a constant collection of feedback from clients, community members, and stakeholders is administered through tablet-based surveys.

2.3 Data Source

The data used for this Canadian Institutes for Health Research funded study came from a sample of 350 youth participants who visited that Vancouver Granville Foundry Centre in 2017.
between July and August, who were all between the ages of 15 and 24. In the primary study, eligible youth, aged 19 and over, followed consenting protocols outlined by the University of British Columbia Ethics Board (REB ID# H17-00127). For youth aged 18 or less, assent protocols, followed by a trained research team member, were used. Participants were recruited in three recruitment periods. Only data from the second and third period were used in this study as this is when the GSS was included in the study survey. The data for this study came from the 204 participants who were recruited in periods two and three when the GSS was interested into the data collection pool. Demographic information collected in the questionnaire included age, gender, ethnicity, self-reported mental health diagnoses. In addition to the GSS, the Kessler Psychological Distress Scale 10 (K10), Patient Health Questionnaire 9 (PHQ-9), and Generalized Anxiety Disorder 7 (GAD-7) were all self-administered.

2.4 Measures

2.4.1 GSS and GSS-I

The GSS and GSS-I (Version 3) were the primary tools of interest for this study and were self-administered along with the K10, PHQ-9, and GAD-7. The GSS contains 23 items which screen for internalizing disorders, externalizing disorders, substance use problems, and crime and violence with the overall tool being referred to as the total disorder screener. In addition, the GSS-I is the sub-screener responsible for measuring internalizing problems and makes up six items of the overall GSS. The GSS was developed from the GAIN to serve three purposes: screening in diverse populations, to determine the level of need across systems, and finally to monitor changes in behavioural disorders overtime. The latest version of the tool (Version 3) varies from previous iterations due to the inclusion of an additional internalizing item, and two externalizing items. The total disorder screener of the tool was reported to have excellent
sensitivity (90%) and specificity (92%) for its total disorder screener when validated against the GAIN which was administered as a clinical assessment. The method in which the GSS is scored is based on how recent an individual has experienced significant problems with particular symptoms. For example, someone who has experienced significant internalizing problems in the past month (maximum score of 3) will score higher on a given item than someone who has not experienced significant internalizing problems at all (minimum score of 0). In order to screen positive on the GSS, a cut-point between 1 and 2 is considered a moderate score on the total disorder screener (GSS) and is recommended for adolescents and adults. For adolescents in outpatient settings a median score of 6 has been observed, with 78% of participants scoring high between 3 and 23. A copy of the questionnaire and scoring guide can be found in Appendix E.

2.4.2 K10

The K10, developed by Kessler in 2003, is a 10-item screening tool used to measure global distress or non-specific mood and anxiety disorders as outlined in the DSM-IV, which was the version of the DSM available at the time of development. As the K10 is a measure of global distress related to non-specific mood and anxiety disorders, its use can be thought of as a trans-diagnostic screening tool similar to the GSS. In the current study, the K10 was used as a reference tool to examine the sensitivity and specificity of the GSS. The justification to use the K10 as a reference tool was based on its similarity to the GSS as a brief screening tool and in its trans-diagnostic approach to measuring distress. The K10 differs from the GSS in that it is a much briefer scale and is primarily a measure of global distress that does not directly assess issues related to externalizing disorders (e.g., problems related to substance use and crime and violence). Although the K10 focuses on assessing symptom severity, it can be used as a screening tool to identify people with suspected clinical conditions. For example, the K10 has
been validated multiple times among adults, across various cut-points, with values ranging between 77% and 86% for sensitivity, and 74% and 83% for specificity.\textsuperscript{80,81} Ideally, a gold standard would have higher values for sensitivity and specificity, and for this reason the K10 will be referred to as a reference tool in the context of this study. Additionally, the brief and non-specific nature of the K10 will inherently pose some limitations in its ability to capture every possible case with a high degree of accuracy.

Scoring the K10 involves adding up the total scores of individual items and using the approved clinical cut points for the tool. In adult populations subjects with a score of 20 or greater are described as having a mild mental disorder, 25 or greater are described as having a moderate mental disorder, and 30 or greater are described as having a severe mental disorder. Moving forward, it is important to consider that these cut points are approved for screening in adults and youth-specific cut points may need to be explored in the future. A copy of the questionnaire and scoring guide can be found in Appendix E.

\textbf{2.4.3 PHQ-9}

The Patient Health Questionnaire (PHQ) is a family of measurements containing the PHQ-9, a nine item screening tool used for the screening, diagnosis, and measurement of depression severity.\textsuperscript{76} The PHQ-9 consists of 9 items, each scored on a 4-point rating scale, with a higher score indicating greater severity of depression. At the time of its development, the PHQ-9 was considered unique because it was half the length of traditional tools used for the measurement of depression and was based on the diagnostic criteria for depression described in the DSM-IV.\textsuperscript{76} The PHQ-9 has demonstrated acceptable validity through testing in multiple adult clinical populations where a clinical assessment was used as the gold standard. Following this validation study the sensitivity and specificity for the PHQ-9 were both reported as 88\%.\textsuperscript{76}
Richardson et al. (2009)\textsuperscript{82} reported that despite the PHQ-9 needing a higher cut-point when screening for depression among adolescents, its validity in terms of sensitivity (89.5\%) and specificity (77.5\%) was similar to findings in adult populations.\textsuperscript{82} For screening purposes, a minimal cut-point of 5 indicates the presence of mild depression. Following this, scores between 10 and 14 indicate moderate depression symptom severity, scores between 15 and 19 indicate moderately severe depression symptoms severity, and scores between 20 and 27 indicate severe depression symptom severity.\textsuperscript{76} A copy of the questionnaire and scoring guide can be found in Appendix E.

2.4.4 GAD-7

The Generalized Anxiety Disorder Scale (GAD-7) is a screening tool that was developed around the diagnostic criteria for anxiety as per the DSM-IV,\textsuperscript{77} similar to the PHQ-9. The original tool consisted of 13 items, nine of which were the DSM-IV criteria for diagnosing GAD-7 in patients and the remaining four items were chosen based on their use in pre-existing screening tools.\textsuperscript{77} The seven items that were chosen for the final version of the tool were those that had the greatest correlation coefficient with the full version containing 13 items. Spitzer et al. (2006)\textsuperscript{77} validated the GAD-7 in a clinical population where the average age was 47.4, but consisted of a range of 18 to 95 years. Young people were part of this sample, but currently there appears to be no study validating the GAD-7 in a sample of youth. When using an optimized cut point of 10, Spitzer et al. (2006) reported that GAD-7 had sensitivity and specificity values of 89\% and 82\% respectively.\textsuperscript{77} For screening purposes, a minimal cut-point of 5 indicates the presence of mild anxiety symptom severity. Following this, scores between 10 and 14 indicate the presence of moderate anxiety symptom severity, and scores between 15 and 21 indicate the
presence of severe anxiety symptom severity. A copy of the questionnaire and scoring guide can be found in Appendix E.

2.5 Analyses

Descriptive statistics were used to characterize the sample (n=201) and groups were divided by self-reported gender identity. The total scores of each screening tool were used to create a Pearson’s correlation coefficient matrix for each pair-wise comparison. The current study used Cronbach’s alpha to measure internal consistency and Cohen’s kappa coefficient to assess agreement. The validity of the GSS as a screening tool was examined by determining the sensitivity and specificity from two-by-two tables when the K10 was used as a reference tool. The cut point chosen for the GSS was 1, and the cut point chosen for the K10 was 20. To optimize the sensitivity and specificity of the GSS, Receiver Operator Characteristic (ROC) curves were generated. Visually, the ROC curves demonstrated the relationship between sensitivity and the false positive rate of the GSS when the K10 was used as a reference tool. To determine optimal cut points for the GSS in the context of this study, Youden’s J statistic was calculated. Area Under the Curve (AUC) values, which provide an estimate of the predictive power of the tool, were also examined from each ROC curve. The examination of the performance of the GSS-I followed a similar process to the GSS, where the K10, PHQ-9, and GAD-7 were used as reference tools to examine the sensitivity and specificity of the GSS-I and generate ROC curves. All statistical analyses were completed using R Studio version 1.1.456, and the syntax that was used to compute the analysis is provided in Appendix B.
2.5.1 **Descriptive Statistics**

Descriptive statistics are essential for the analysis of any sample as they help describe the characteristics of a sample with measures of central tendency such as mean, median, and mode. In this study the mean and median age of participants were calculated, which are simply the average age and the middle age. Other variables such as ethnicity, educational status, and self-reported diagnoses will be summarized with counts and proportions by self-reported gender identity.

2.5.2 **Pearson’s Correlation Coefficient and Correlational Matrix**

Pearson’s correlation coefficient (r) is a value between -1 and 1 that describes the strength of a linear relationship between two variables. Coefficient values less than 0 indicate negative relationships, and coefficient values greater than 0 represent positive relationships. The assumptions of the Pearson correlation test are as follows: first, both variables are continuous; second, the relationship is linear; third, all values are independent of another; and finally all values are normally distributed. If these assumptions are not met, then Spearman’s Rank-Order Correlation may be used instead as it is a non-parametric test with fewer assumptions required. Preliminary tests of normality and linearity revealed that the variables met the assumptions of Pearson’s correlation coefficient, and these results are provided in Appendix C and D respectively. A correlation matrix for every pair-wise comparison was used to summarize the correlation coefficients and their respective 95% confidence intervals. In this study, it was expected that the relationships between the screening tools will generally be positive with varying strengths of correlation across screening tool comparisons. The GSS and GSS-I have items which are scored based on how recently an individual experienced significant problems with that particular item, whereas the K10, PHQ-9, and GAD-7 have items which are scored...
based on how severe the individual is experiencing symptoms. Given this, it is hypothesized that the GSS and GSS-I will not be strongly correlated with the K10, PHQ-9, and GAD-7 given that the tools are structured slightly differently.

2.5.3 Cronbach’s Alpha

Cronbach’s alpha is one of the most common statistics used in the objective measurement of reliability. More specifically, Cronbach’s alpha is a measure of internal consistency and is used to explain the level of which items in a test measure the same construct.\textsuperscript{84} The range of Cronbach’s alpha is between 0 and 1, where values between 0.70 and 0.95 are cited as acceptable values. In contrast with Pearson’s correlation coefficient, a higher alpha score is not necessarily a greater indication of a more reliable test. A very high Cronbach’s alpha can indicate that some items within a test are redundant or repeated measures, as opposed to being reliable measures of the same construct. For this reason it has been suggested that the maximum acceptable alpha score be 0.90 for the general population.\textsuperscript{84} One of the primary assumptions of Cronbach’s alpha is that all items are unidimensional. For an item to be unidimensional, instead of multidimensional, it must only measure a single construct. If this assumption is violated, along with other assumptions for Cronbach’s alpha, it is likely that the resulting Cronbach’s alpha score will be lower.\textsuperscript{91}

2.5.4 Cohen’s Kappa Statistic

When measurement occurs by two different individuals, Cohen’s Kappa Statistic is used to measure the rate at which the observers agreed or disagreed based on chance alone.\textsuperscript{83} This is also known as inter-rater reliability, and provides an estimate of how likely the observed agreement was due to chance. In the context of this study, Cohen’s Kappa is being used to
determine the level of agreement between the same individual rating two different screening tools. For this reason, the interpretation of Cohen’s Kappa in this study will not be as a measure of inter-rater agreement but simply as a measure of agreement adjusted for chance.

2.5.5 Sensitivity and Specificity

Sensitivity and specificity represent two of the primary markers of screening test validity in measurement. Sensitivity refers to the ability of a test to correctly identify positive cases who actually have the outcome of interest. Specificity refers to the opposite as sensitivity, which is the ability of a test to correctly rule out participants without the outcome of interest. The terms “true positive” and “true negative” can also be used to explain the occurrences described by sensitivity and specificity respectively. In measurement, a false positive occurs when a test detects a positive case when they do not have the outcome of interest. Conversely, a false negative occurs when a test declares a case negative when they actually have the outcome of interest. Sensitivity and specificity are measured as percentages, with 100% in both measures representing a perfectly accurate test. In reality, there is no perfectly accurate test; however, some tests do have excellent sensitivity and specificity. As previously discussed, screening tools which have been determined to be valid with excellent sensitivity and specificity can be used as gold standards to validate other screening tools. In this study, sensitivity and specificity were examined using two-by-two tables for the GSS and GSS-I with their respective reference tools.

2.5.6 Positive and Negative Predictive Values

Positive Predictive Values (PPV) and Negative Predictive Values (NPV) are measures which represent the proportions of true positives and true negatives. PPV represents the
proportion of cases who are true positives among all cases who tested positive, whereas NPV represents the proportion of cases who are true negatives among all cases who tested negative. Predictive values are also measures of screening test validity and in optimal cases would be 100% but this is rare for most screening tools. It is important to note that the prevalence of outcomes of interest can affect PPV and NPV. Generally, the rarer an outcome is (low prevalence in the population) then the lower PPV and higher NPV will be. In a population where the prevalence of an outcome is very high, it is likely that the majority of cases will test as true positive. Conversely, if the majority of people in a population have the outcome then there will simply be fewer people in the population who do not have outcome and lower the NPV. PPV and NPV along with sensitivity and specificity help to provide a thorough understanding of the psychometric properties of a screening test. Before stating the results of this study, it is important to state the background prevalence of anxiety and mood disorders in youth to indicate whether or not the outcome is rare. As previously stated, in a CCHS survey of 5000 Canadian youth 10% reported they had depression and 12% reported they had anxiety. Using this data, the assumption will be made that outcomes of mood and anxiety disorders in youth are not rare. This same assumption will be applied to mental health concerns in youth in general as depression and anxiety were not the only disorders measured by the GSS.

2.5.7 ROC Curves

Receiver Operator Characteristic (ROC) curves allow for graphical representations of the relationship between sensitivity and specificity of a test. An ROC curve can be constructed by graphing the sensitivity of a test on one axis by the False Positive Rate (FPR) on the other axis. The FPR is calculated by subtracting values of specificity of 1 and represents the inverse of specificity. ROC curves help determine the point at which sensitivity and specificity are optimal
for a tool, and then be used to examine the optimal cut points for a tool. A cut point is a statistic that is used when interpreting the scores of a clinical test and helps flag participants if they are at risk of developing the outcome of interest. For example, in the GSS total disorder screener a minimum cut point of 1 is recommended to indicate the risk of a moderate disorder. Visually, on a ROC curve the optimal cut point should be the upper-left most point before the curve plateaus. One method in which optimal cut points can be determined involves the calculation of Youden’s J Statistic.85 Simply, Youden’s J can be calculated by adding the sensitivity and specificity of a screening tool across various cut points then subtracting 1 (to create a value between 0 and 1).85 Following this, the highest value of Youden’s J across all cut points would be the optimal cut point.

2.5.8 AUC

The Area Under the Curve (AUC) is a value between 0 and 1 which represents the performance of a test in its ability to classify cases, based on the area under the ROC curve.95 Essentially, the higher the AUC value the greater power the test has to distinguish between cases with the outcome and cases without the outcome. In an ideal ROC curve where the AUC value is 1, at virtually any given point on the curve the rate of true positives should be 100%.

2.5.9 Missing Data Analysis

Of the 204 participants, three (<1%) did not provide valid responses to the majority of the survey questions and were removed from the dataset. It was assumed that these participants were missing completely at random and their removal would not impact the sample. Among the remaining 201 participants, the amount of missing data was below 3% for all items and the missing values appeared to be missing completely at random (Little's MCAR test: Chi-Square =
223, DF = 604, p > .05). To maintain sample size, missing responses for the 201 participants were imputed using single imputation with the EM algorithm in SPSS Version 25.
Chapter 3: Results

3.1 Descriptive Statistics

The results from the descriptive analysis of the sample are provided in Table 1. In this sample the average age of participants was 21.3 years with a standard deviation of 1.8 years, and a range between 17 and 24. The median age, which was 21, was approximately equal to the average age and the inter-quaternary range (IQR) was 3. The average age of women and men in the study were approximately equal, with the average age of non-binary participants being approximately three years less than the average age of women and men.

The proportion of gender is essentially equal for women and men at approximately 44% each, with the remainder of participants identifying outside gender of the binary at approximately 12%.

The ethnicity of participants was collapsed into four categories which reflected the most prevalent ethnic characteristics of the sample: Caucasian, Indigenous (First Nations, Metis, and/or Inuit), Multiple Ethnicities, and Other. Caucasian participants represented over half the sample with 51.7% of participants identifying their ethnicity as Caucasian. Indigenous groups represented a significant minority in this sample of youth with 14.4% of participants identify as Indigenous. Individuals who belonged to more than one of the provided ethnic groups were categorized as having “Multiple Ethnicities” and represented 14.4% of the sample. Participants who identified any other ethnicity made up the remainder of the population and represented 18.4% of the sample. In terms of educational attainment, the vast majority of participants were either in the process of completing high school or had attained a high school diploma and represented 69.2% of the population. A significant minority of participants who initiated post-secondary education (college or university) at the time of this study and represented 24.4% of this sample, and 5.5% of participants had completed post-secondary education at the time of this
study. When participants were asked to self-report their diagnoses, they were able to select any of the listed options that applied to their current circumstance. In most categories, women self-reported diagnoses in greater amounts than men and non-binary participants. Anxiety (50.7% of women, 34.6% of men, 14.7% of non-binary participants) and depression (49.2% of women, 35.7% of men, 15.1% of non-binary participants) were the most endorsed diagnoses across all gender identities. Following this, other diagnoses that were not listed, post-traumatic disorder, and bipolar disorder were significant minorities in the sample. Self-reported diagnoses for schizophrenia and schizoaffective disorder were the rarest.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Women</th>
<th>Men</th>
<th>Non-Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Age (years, SD)</strong></td>
<td>21.4 (1.8)</td>
<td>21.2 (2.0)</td>
<td>18.4 (2.1)</td>
</tr>
<tr>
<td><strong>Gender (n, %)</strong></td>
<td>88 (43.8)</td>
<td>89 (44.3)</td>
<td>24 (11.9)</td>
</tr>
<tr>
<td><strong>Ethnicity (n, %)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>47 (45.2)</td>
<td>44 (42.3)</td>
<td>13 (12.5)</td>
</tr>
<tr>
<td>Indigenous (First Nations, Metis, and/or Inuit)</td>
<td>11 (37.9)</td>
<td>17 (58.6)</td>
<td>1 (3.5)</td>
</tr>
<tr>
<td>Multiple Ethnicities</td>
<td>11 (37.9)</td>
<td>12 (41.4)</td>
<td>6 (20.7)</td>
</tr>
<tr>
<td>Other</td>
<td>19 (51.4)</td>
<td>16 (43.2)</td>
<td>2 (5.4)</td>
</tr>
<tr>
<td><strong>Education (n, %)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some High School</td>
<td>29 (35.8)</td>
<td>40 (49.4)</td>
<td>12 (14.8)</td>
</tr>
<tr>
<td>High School Diploma</td>
<td>24 (41.4)</td>
<td>29 (50.0)</td>
<td>5 (8.6)</td>
</tr>
</tbody>
</table>
Table 1 - Descriptive statistics and demographic characteristics of the study sample, including age, gender, ethnicity, education, and self-reported diagnoses (n=201).

<table>
<thead>
<tr>
<th>Educational Level</th>
<th>Some College or University</th>
<th>Post-Secondary Diploma/Degree</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29 (59.2)</td>
<td>15 (30.6)</td>
<td>5 (10.2)</td>
</tr>
<tr>
<td></td>
<td>6 (54.5)</td>
<td>5 (45.5)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Self-Reported Diagnoses (n, %)</th>
<th>Anxiety</th>
<th>Bipolar Disorder</th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69 (50.7)</td>
<td>23 (59.0)</td>
<td>62 (49.2)</td>
</tr>
<tr>
<td></td>
<td>47 (34.6)</td>
<td>12 (30.7)</td>
<td>45 (35.7)</td>
</tr>
<tr>
<td></td>
<td>20 (14.7)</td>
<td>4 (10.3)</td>
<td>19 (15.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Traumatic Stress Disorder</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>34 (47.9)</td>
<td>22 (31.0)</td>
<td>15 (21.1)</td>
</tr>
<tr>
<td>Schizoaffective Disorder</td>
<td>1 (12.5)</td>
<td>6 (75.0)</td>
<td>1 (12.5)</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>1 (8.3)</td>
<td>10 (83.4)</td>
<td>1 (8.3)</td>
</tr>
<tr>
<td>Other</td>
<td>32 (41.0)</td>
<td>35 (44.9)</td>
<td>11 (14.1)</td>
</tr>
<tr>
<td>Multiple Diagnoses</td>
<td>65 (50.4)</td>
<td>43 (33.3)</td>
<td>21 (16.3)</td>
</tr>
</tbody>
</table>

3.2 **Bivariate Correlation Statistics**

To determine the level of correlation between the total scores of each individual screening tool, the Pearson’s (r) correlational matrix in Table 2 was tabulated. Overall, the GSS was weakly correlated with the K10, the PHQ-9, and the GAD-7 with values between 0.32 and 0.39 and moderately correlated with the GSS-I with a value of 0.71. The K10 was strongly correlated with the PHQ-9 and GAD-7 with scores of 0.84 and 0.82 respectively, and moderately
correlated with the GSS-I with a value of 0.56. The PHQ-9 and the GAD-7 are different tools which measure depression and anxiety respectively; however, the two tools were strongly correlated with each other with a score of 0.80. Finally, the GSS-I demonstrated a moderate correlation with the PHQ-9 with a score of 0.56 and a weak to moderate correlation with the GAD-7 with a score of 0.47. Pearson’s Correlational Test revealed that all relationships were statistically significant with p-values less than 0.01.

<table>
<thead>
<tr>
<th>Screening Tool</th>
<th>GSS Total</th>
<th>K10 Total</th>
<th>PHQ-9 Total</th>
<th>GAD-7 Total</th>
<th>GSS-I Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS Total</td>
<td>1</td>
<td>0.37</td>
<td>0.39 (0.26-0.50)</td>
<td>0.32 (0.19-0.44)</td>
<td>0.71 (0.63-0.77)</td>
</tr>
<tr>
<td>K10 Total</td>
<td>0.37 (0.24-0.48)</td>
<td>1</td>
<td>0.84 (0.80-0.88)</td>
<td>0.82 (0.77-0.86)</td>
<td>0.56 (0.46-0.65)</td>
</tr>
<tr>
<td>PHQ-9 Total</td>
<td>0.39 (0.26-0.50)</td>
<td>0.84 (0.80-0.88)</td>
<td>1</td>
<td>0.79 (0.74-0.84)</td>
<td>0.56 (0.45-0.64)</td>
</tr>
<tr>
<td>GAD-7 Total</td>
<td>0.32 (0.19-0.44)</td>
<td>0.82 (0.77-0.86)</td>
<td>0.79 (0.74-0.84)</td>
<td>1</td>
<td>0.47 (0.36-0.58)</td>
</tr>
<tr>
<td>GSS-I Total</td>
<td>0.71 (0.63-0.77)</td>
<td>0.56 (0.46-0.65)</td>
<td>0.56 (0.45-0.64)</td>
<td>0.47 (0.36-0.58)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 - Pearson’s Correlational coefficient (r), measuring the degree of linearity between two continuous variables for each screening tool used in the current study and respective 95% confidence intervals.
3.3 Cronbach’s Alpha

Analysis of Cronbach’s alpha indicated that each tool has generally acceptable internal consistency with the GSS having an alpha of 0.91, the GSS-I having 0.83, the K10 having 0.94, the PHQ-9 having 0.91 and the GAD-7 having 0.91.

3.4 Cohen’s Kappa Coefficient

In Tables 3, 4, and 5, the results of Cohen’s Kappa are provided for the GSS, the GSS-I, and the remaining screening tools respectively. All possible comparisons were tabulated, similar to the correlational matrix above, for the purpose of understanding the level of agreement adjusted for chance between all the screening tools. Agreement was poor for most comparisons with the GSS, with scores ranging from 0.091 to 0.20, and the highest value of 0.61 occurring between the GSS and the GSS-I. Agreement for the GSS-I with the K10, PHQ-9, and GAD-7 was slightly better with values ranging from 0.25 to 0.39. The agreement between the K10, PHQ-9, and GAD-7 was the greatest with values ranging from 0.44 to 0.47.

<table>
<thead>
<tr>
<th>Interrater Comparison</th>
<th>Cohen’s Kappa Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS and K10</td>
<td>0.14</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>GSS and GSS-I</td>
<td>0.61</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>GSS and PHQ-9</td>
<td>0.20</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>GSS and GAD-7</td>
<td>0.091</td>
<td>0.0678</td>
</tr>
</tbody>
</table>

Table 3 - Values for Cohen’s Kappa Coefficient for the GSS with all other tools and corresponding p-values for tests of significance.
### Interrater Comparison

<table>
<thead>
<tr>
<th>Interrater Comparison</th>
<th>Cohen’s Kappa Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS-I and K10</td>
<td>0.25</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>GSS-I and PHQ-9</td>
<td>0.39</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>GSS-I and GAD-7</td>
<td>0.39</td>
<td>p&lt;0.01</td>
</tr>
</tbody>
</table>

Table 4 - Values for Cohen’s Kappa Coefficient for the GSS-I with all other tools and corresponding p-values for tests of significance.

<table>
<thead>
<tr>
<th>Interrater Comparison</th>
<th>Cohen’s Kappa Coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>K10 and PHQ-9</td>
<td>0.47</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>K10 and GAD-7</td>
<td>0.44</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td>PHQ-9 and GAD-7</td>
<td>0.45</td>
<td>p&lt;0.01</td>
</tr>
</tbody>
</table>

Table 5 - Values for Cohen’s Kappa Coefficient for the K10, PHQ-9, and GAD-7 with all other tools and corresponding p-values for tests of significance.

### 3.5 Analysis of the GSS Total Disorder Screener Using the K10 as a Reference Tool

#### 3.5.1 GSS Sensitivity and Specificity

Sensitivity and specificity values observed when the GSS was referenced with the K10 are summarized in Table 6, and the resulting Receiver Operator Characteristic (ROC) curve is provided in Figure 2. When using the recommended cut-point of 1 for the GSS, the sensitivity and specificity were 100.0% and 9.6% respectively. Using the ROC curve and Youden’s J statistic, an optimal cut-point of 7 was identified. At this point the observed sensitivity was 90.6% and the specificity was 42.3%. A cut point of 10 is required to achieve a minimum specificity of 50%. At this point, the observed sensitivity and specificity were 76.3% and 51.9%.
respectively. For the analysis of the GSS with the K10, the AUC value was 0.67, indicating that the predictive capabilities of the GSS when validated with the K10 are moderate.

<table>
<thead>
<tr>
<th>Cut-point</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>9.6</td>
<td>0.67</td>
</tr>
<tr>
<td>7</td>
<td>96.0</td>
<td>40.4</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>90.6</td>
<td>42.3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>87.9</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>76.3</td>
<td>51.9</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 - Values for sensitivity, specificity, and AUC for the GSS across optimal cut-points determined through ROC curves.

3.5.2 GSS PPV and NPV

The PPV and NPV across various cut points for the GSS when referenced with the K10 are presented in Table 7. Generally, the PPV seems to plateau around the cut points between 7 and 10 but the NPV steadily declines with each incremental cut point.

<table>
<thead>
<tr>
<th>Cut-point</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>76.0</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>82.2</td>
<td>77.8</td>
</tr>
<tr>
<td>8</td>
<td>81.8</td>
<td>61.1</td>
</tr>
<tr>
<td>9</td>
<td>81.9</td>
<td>56.1</td>
</tr>
<tr>
<td>10</td>
<td>82.2</td>
<td>45.0</td>
</tr>
</tbody>
</table>

Table 7 - Positive Predictive Values (PPV) and Negative Predictive Values (NPV) NPV across various cut points for the GSS when referenced with the K10.
3.5.3 GSS ROC Curve

The ROC curve for the analysis of the GSS with the K10 as a reference tool is provided in Figure 4.1. It is clear from observing the shape of the curve that 7 is the optimal cut point for the curve, consistent with all the findings in this study. At point 7 the curve is at its upper-left most point before beginning to plateau, indicating this is the point at which sensitivity and specificity are optimal.

![ROC Curve](image)

Figure 2 - ROC Curve for the analysis the GSS with the K10 as a reference tool with various cut points identified.
3.6 Analysis of the GSS-I using the K10, PHQ-9, and GAD-7 as Reference Tools

3.6.1 GSS-I Sensitivity and Specificity

The results of the analysis of the GSS-I using the K10, PHQ-9, and GAD-7 as reference tools are summarized in Table 8. An optimal cut-point of 4 using Youden’s J statistic was determined for the GSS-I consistently throughout analysis.

When referenced with the K10, the sensitivity and specificity of the GSS-I when using the recommended cut point of 1 were 99.3% and 19.2% respectively. At the optimal cut point of 4 the sensitivity and specificity were 91.3% and 48.1% respectively. The AUC value observed when the GSS-I was referenced with the K10 was 0.71, indicating that the GSS-I has moderate predictivity capabilities when the K10 was used as a reference tool.

When referenced with the PHQ-9, the sensitivity and specificity of the GSS-I when using the recommended cut point of 1 were 97.8% and 33.3% respectively. At the optimal cut point of 4 the sensitivity and specificity were 86.7% and 66.7% respectively. The AUC value observed when the GSS-I was referenced with the PHQ-9 was 0.78, indicating that the GSS-I has moderate predictivity capabilities when the PHQ-9 was used as a reference tool.

When referenced with the GAD-7, the sensitivity and specificity of the GSS-I when using the recommended cut point of 1 were 98.3% and 30.8% respectively. At the optimal cut point of 4 the sensitivity and specificity were 86.3% and 53.8% respectively. The AUC value observed when the GSS-I was referenced with the GAD-7 was 0.72, indicating that the GSS-I has moderate predictivity capabilities when the GAD-7 was used as a reference tool.
<table>
<thead>
<tr>
<th>Cut-point</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the K10</td>
<td>1</td>
<td>99.3</td>
<td>19.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>99.3</td>
<td>21.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>97.3</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>91.3</td>
<td>48.1</td>
</tr>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the PHQ-9</td>
<td>1</td>
<td>97.8</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>97.8</td>
<td>38.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>94.4</td>
<td>47.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>86.7</td>
<td>66.7</td>
</tr>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the GAD-7</td>
<td>1</td>
<td>98.3</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>98.3</td>
<td>34.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>94.3</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>86.3</td>
<td>53.8</td>
</tr>
</tbody>
</table>

Table 8 - Values for sensitivity, specificity, and AUC for the GSS-I across optimal cut-points determined through ROC curves.

3.6.2 GSS-I PPV and NPV

The PPV and NPV for various cut points of the GSS-I referenced with the K10, PHQ-9, and GAD-7 are provided in Table 9. The trends for PPV and NPV across the three different screening tools that were used to validate the GSS-I are similar. Youden’s J statistic revealed that the optimal cut point for the GSS-I was 4 for all three sets of analyses. For the GSS-I it appears that values for PPV increase and the values for NPV decrease with each successive cut point.
<table>
<thead>
<tr>
<th>Cut-point</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the K10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>77.9</td>
<td>90.9</td>
</tr>
<tr>
<td>2</td>
<td>78.3</td>
<td>91.7</td>
</tr>
<tr>
<td>3</td>
<td>80.1</td>
<td>80.0</td>
</tr>
<tr>
<td>4</td>
<td>83.4</td>
<td>65.8</td>
</tr>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the PHQ-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>92.6</td>
<td>63.6</td>
</tr>
<tr>
<td>2</td>
<td>93.1</td>
<td>66.7</td>
</tr>
<tr>
<td>3</td>
<td>93.9</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>95.7</td>
<td>36.8</td>
</tr>
<tr>
<td><strong>GSS-I</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validated with the GAD-7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>90.5</td>
<td>72.7</td>
</tr>
<tr>
<td>2</td>
<td>91.0</td>
<td>75.0</td>
</tr>
<tr>
<td>3</td>
<td>91.2</td>
<td>50.0</td>
</tr>
<tr>
<td>4</td>
<td>92.6</td>
<td>36.8</td>
</tr>
</tbody>
</table>

Table 9 - Positive Predictive Values (PPV) and Negative Predictive Values (NPV) NPV across various cut points for the GSS-I when validated with the K10, PHQ-9, and GAD-7.

### 3.6.3 GSS-I ROC Curves

The ROC curves for the analysis of the GSS-I with the K10, PHQ-9, and GAD-7 are presented in Figures 3, 4, and 5 respectively. All three curves display similar behaviour of the GSS across the three different reference tools. Based on observation it appears that the optimal cut point for the GSS-I occurs at 4, which is consistent with result of Youden’s J statistic.
Figure 3 - ROC Curve for the analysis of the GSS-I with the K10 as a reference tool with various cut points identified.

Figure 4 - ROC Curve for the analysis of the GSS-I with the PHQ-9 as a reference tool with various cut points identified.
Figure 5 - ROC Curve for the analysis of the GSS-I with the GAD-7 as a reference tool with various cut points identified.
Chapter 4: Discussion

4.1 Summary of Findings

The purpose of this study was to contribute to the literature on the validity of the GSS and the GSS-I by examining their sensitivity and specificity in an integrated youth health service. Sensitivity and specificity of the GSS were assessed using the K10 as a reference tool. The results indicated that the GSS had acceptable sensitivity; however, this appears to be accompanied by a steep trade off in specificity, even when an optimized cut-point was used. To better understand this compromise, it is important to first consider how the GSS is scored. In order to score the minimum of 1 point on a GSS item, a participant must have indicated they were experiencing significant problems with an item in the past year. When interpreting scores, the GAIN manual recommends that anyone whose total score is between 1 to 2 out of 23 may benefit from a full assessment, and estimates that 50% of participants who scored in this range would receive a diagnosis following a clinical assessment. Given this, it seems reasonable to conclude that the GSS is designed to flag as many participants as possible (or approximately 98% of this sample) who present any past year problems related to a mental health concern. Given this aspect of how the GSS is scored, the optimal cut points that were determined as part of this study should be assessed to determine if they can mitigate this issue.

When a cut point of 1 was used for the GSS, approximately 98% of the sample was reported as “moderate” based on their symptoms. An optimal cut point as determined by the use of ROC curves and Youden’s J statistic was 7, and this lowered the number of individuals flagged as moderate down to 87%, or a difference of 22 participants. Even with a cut point of 7, the specificity of the GSS was not very acceptable (40.4%) and the aforementioned 22 individuals who subsequently screened as negative may be false negatives. At this point, it is important to weigh the potential benefits and consequences of using an optimal cut point. That is
to say, although most individuals will benefit from a second, complete assessment, these assessments may be time consuming and exhaust resources. From a clinical standpoint, having to perform 22 less clinical assessments may be much more cost-effective. On the other hand, it may not be of the best interest to the patients if there is not a high degree of certainty in their negative results. Similar patterns of results can be observed with the GSS-I, where even with an optimal cut point the specificity is still not perfect. In summary, although this study determined an optimal cut point for the GSS and GSS-I, it may not be entirely useful given the inability of the GSS and GSS-I to accurately discriminate between true and false negatives.

With regards to the GSS-I, this study provided new evidence that contributes to literature of the on-going the validation of the GSS. This appears to be one of the first studies to assess the sensitivity and specificity of the GSS-I independent of the rest of the tool. The findings for the GSS-I were similar to those of the GSS, in that both tools will have acceptable sensitivity and low specificity when using the recommended cut-point of 1. As with the GSS, the low sensitivity of the GSS-I can be addressed by ensuring that valid second stage assessments are easily accessed by youth who initially screen positive on the GSS-I. It is clear that the GSS is not strongly correlated with the K10; however, from the results it appears that the GSS-I and K-10 have a greater overall fit based on a higher Pearson correlation coefficient and AUC value. Given that the GSS-I is shorter than both the PHQ-9 and GAD-7 yet is technically a measure of both depression and anxiety, it may be suitable to use the GSS-I to screen for anxiety and depression through a trans-diagnostic approach. This convenience can be beneficial for youth with emerging mental illnesses, such as those who are in Stage 0 of the proposed clinical staging model, but comes at the cost of not providing clinicians and youth with separate assessments for anxiety and depression, which may complicate the discussion of test results with clients. As previously discussed in the background of this study, there have been efforts made by previous researchers
to create a tri-partite model for the diagnosis of anxiety and depression.\textsuperscript{14} The GSS-I can be used to complement a tri-partite model by first determining if someone is experiencing significant internalizing problems and then following up with a secondary assessment to determine if their diagnosis is anxiety-specific, depression-specific, or a combination of the two. One of the objectives of this study was to study the GSS-I and provide recommendations for its potential to be used independently of the rest of the tool. Based on the results of this study, clinicians and researchers should proceed with caution as further evidence is required to determine if the GSS-I is acceptable for screening in this setting. Overall, the measurement strategy should not be driven by efficiency alone, but rather should be patient-centred and based on the strongest principles of measurement science.\textsuperscript{99,100}

4.2 Over-Diagnosis and Over-Screening

One of the controversial aspects of screening tools with high sensitivity and low specificity is a phenomena known as “over-screening”.\textsuperscript{101} In oncology, over-screening is seen as an issue for detecting breast cancer, where the second stage may involve an invasive procedure such as a mammogram or biopsy.\textsuperscript{102} Conversely, the second stage of screening for mental illnesses typically involves a clinical assessment to establish a formal diagnosis. While the use of the GSS may lead to a high number of false positives initially, the consequences associated with a false positive (“over-screening”) are likely minimal in this setting where secondary assessment (e.g., via completion of the GAIN in a clinical assessment) is readily available. This finding aligns with the recommendations of Lalkhen et al. (2008)\textsuperscript{93} who discussed the implications of screening tools with high sensitivity and low specificity. They recommended that the use of a highly specific tool during a secondary assessment serves to limit the over-diagnosis of mental disorders when screening programs are implemented.\textsuperscript{93}
4.3 Support for a Clinical Staging Model

As previously described, a recent paradigm shift in psychiatry has urged researchers and clinicians to stage mental illnesses based on risk, the presence of symptoms, and the onset of psychiatric episodes. This study supports the further development of clinical staging models as the GSS provides useful information on the duration and progression of an illness. The GSS contrasts the other screening tools used in the study due to the measurement of symptoms based on how recently they were experienced. If someone were to complete the K10 it would provide a researcher or a clinician an understanding of their current experience with distress and assess severity. In contrast, the GSS assesses how recent the symptoms have been experienced within the past year and may align well with a clinical staging model. The basis of a clinical staging model is to describe the progression of a mental illness from youth into adulthood, creating discrete categories or stages that one could be identified within, and explaining what actions or interventions are appropriate at each stage. Overall, the GSS may provide more useful information than a tool such as the K10 to support the development of clinical staging models due to the collection of data that is based on how recent someone has experienced symptoms.

4.4 Provincial and Pan-Canadian Recommendations for Standardized Use of the GSS

The implementation process for standardizing the use of a screening tool across a country is challenging. Gotham et al. (2008) implemented the GAIN to be used state-wide in Missouri and explained that there are many key factors involved in successful implementation. Among these, the authors stated that the cooperation of all individuals from the client and service level, up to the governmental level was essential. As previously discussed, the GAIN and GSS also have positive support in Canada from the provinces of Ontario and Alberta. Following this study, there are a few considerations that the province of British Columbia and other Canadian
provinces can look towards for implementing the GAIN or GSS. The first thing provinces should consider is what exactly will be the purpose of the screening program. Using the GSS as a widespread screening tool will result in the measurement of four broad categories of mental and behavioural health as opposed to screening for specific illnesses. For example, if a screening program was to use the PHQ-9 or GAD-7 then the outcome of screening would be identifying persons in the population who may have depression or anxiety. Instead, a screening program could opt to use a tool like the GSS or the K10 to broaden its scope. In this instance, the strength of using the GSS over the K10 would be that the GSS also measures substance use and crime and violence problems as well. Despite the fact that the GSS is a slightly longer tool than the K10, the difference in time it takes to administer the tool should be negligible, but this may need to be tested in the future to confirm. Although the results of this study are limited, the GSS may still be a suitable choice as a brief screening tool which could be paired with a secondary assessment.

In order to implement a screening program across a province then it must be established what the follow-up measures will involve. Unless the first stage of screening is a clinical assessment, it is not likely that a screening program can only have one stage. The purpose of screening is cast a wide net to determine a subset of the population that is at risk of developing the outcome of interest. This serves to decrease the overall pool of participants and to ensure that only those that truly have the outcome(s) are diagnosed. For this reason, it is important that researchers consider what secondary assessments they want to use and where they will take place. The benefit of using the GSS as a screening tool is that it readily shuttles into a second stage of screening using the GAIN as a clinical assessment. The GSS was developed from the items in the GAIN and by using them in the same screening program researchers and clinicians can ensure a more congruent process. If the province of British Columbia was to develop a provincial screening program using the GSS and the GAIN, then it must also consider if they
want to use an integrated model like Foundry. A true strength of an integrated service like Foundry is having secondary assessments readily available and the province may want to consider consulting Foundry to learn more about this model. The next paragraph will build on these suggestions and discuss why integrated health services are important for screening youth.

For any province considering the development of a screening program, it must be considered where the program will take place. British Columbia may be in good a position to start using the GSS across the province by continuing the efforts made by Foundry. Foundry has eight sites in the province of British Columbia and currently uses the GSS to screen clients who are accessing their services. Integrated youth health centres are an ideal location for screening as discussed in background of this study which demonstrated that access to mental health care is decentralized. For British Columbia to begin using the GSS across the province it may want to first acknowledge the excellent work that Foundry has done in providing integrated health services to youth all over the province. The province should determine if there would be any overlap in clinical goals for screening programs and ensure there is no redundancy. For example, if the majority of Foundry clients are made up of youth from the community and clinical populations, perhaps the province should focus on screening youth in schools and juvenile detention centres, since they may have greater access to these areas. The background of this study provided information on screening youth in schools and how schools may be well suited to capture a different demographic of youth, such as those under the age of 18. More work is required to determine if the GSS is suitable to screen youth in schools and this may be an area of research that the province could establish. Overall, the location of screening would be important to establish early on so the province could complement the great work already made by Foundry.

Following this study which aimed to determine the suitability of the GSS as a screening tool in an integrated youth health service, it is encouraged that British Columbia proceed with
caution with implementing it as a provincial screening tool. The provinces of Alberta and Ontario have endorsed the use of the GSS as a tool for screening youth and adults for mental illnesses; however, the current research is limited, and a conclusive recommendation cannot be made at the time of this study. In the next section, recommendations for future research will be made to potentially move this area of interest for the GSS forward.

4.5 Future Directions of Research

The primary goals of this study were to provide the results of a detailed search on current GAIN and GSS literature and then to examine the sensitivity and specificity of the GSS and GSS-I for use in an integrated youth health service. After summarizing the results of the literature search on the GSS, there appeared to be very few studies which validated the tool. Although the literature is limited in this regard, three of the four GSS validation studies were focused on adolescent populations which is helpful given the context of the current study. The identifiable gap in the literature this study hopes to address is clear, and additional studies are required to replicate these findings among youth to further support the validation of GSS as a screening tool for integrated youth health services. For future directions of research, there should be important considerations made towards the reference tool chosen to validate the GSS, as the current study experienced limited results with the reference tools chosen.

In this study, it was clear that the GSS was poorly correlated with the selected reference tools. In contrast, the K10 was strongly correlated with the PHQ-9 and GAD-7, and the PHQ-9 and GAD-7 were strongly correlated with each other. This may have been due to the GSS having items which measure how recent someone experienced their symptoms, as opposed to measuring the current severity of someone’s symptoms. In the limitations section of this study, a detailed explanation about why the K10 cannot be thought of as a gold standard will be provided. Future
research should remain consistent with the initial work of Dennis et al. (2006),\textsuperscript{47} and only use the GAIN or another clinical assessment as a gold standard. Despite not being strongly correlated with the reference tools, the GSS still has redeeming qualities as a brief tool with a trans-diagnostic approach that is also cost-effective and easy to administer.

The second research consideration arising from this study is to encourage the continued implementation of screening programs across Canada to study client satisfaction and monitor trends in mental illness. In this section the potential consequences of over-screening were identified, but it is possible to minimize these risks in the context of an integrated youth health service. Given this, the implementation of more mental health screening programs in Canada may also rely on the development of more integrated youth health services. Although Foundry would be an expert source to contact for this work, the onus cannot be placed entirely on them and requires greater Pan-Canadian collaboration. The impacts of greater screening programs will likely happen at a broad, population-based level. As theorized by Dowdy et al. (2010), it is likely that a greater focus on screening and proactive approaches could help transform the current state of mental health care.\textsuperscript{37} By screening and identifying individuals at risk of a mental illness early it is possible that society will begin seeing reductions in severe disability resulting from untreated mental illness. If more screening programs are to be implemented, they should be studied closely to not only monitor trends in the morbidity of mental illness, but also determine if clients are satisfied with the services they provide. As discussed in the previous section, this recommendation for research will be particularly important for British Columbia if they choose to use the GSS or another screening tool provincially.
4.6 Limitations

The current study has several limitations that may have impacted the results. To summarize this section, the limitations that will be addressed are the study’s sample size, the use of the K10 as a reference tool, and measurement error.

First, the sample size for this study is much smaller relative to other studies examining the performance of the GSS. Having a larger sample could provide greater statistical power and allow researchers to draw more definitive conclusions across a broad range of youth populations. In general, this concept is known as external validity, and as previously discussed it one of the most common types of validity along with internal validity. Even though a larger study sample could have benefitted this study, the presence of statistically significant results from the Pearson’s r, Cronbach’s alpha, and Cohen’s kappa indicated that there was a sample size large enough to achieve significance. Future studies which aim to examine the validity of the GSS should also consider drawing their sample from multiple sites to study the differences of the tool across demographics.

Following this study and interpreting the results, the use of the K10 as a reference tool that is equivalent to a traditional gold standard was not justified. It is important to note that the term “gold standard” relies on the tool or measure in question to be very accurate and reliable. When validating a tool to determine its sensitivity and specificity, the selected gold standard should be able to determine all the true positive cases and true negative cases. Previous studies which validated the K10 demonstrated that its sensitivity and specificity were not perfect, and for this reason the use of the term “gold standard” was not appropriate. For the purpose of this study, the K10 was used as a reference to compare the cases identified by the GSS. In this case, the term “reference tool” was more appropriate due to limitations in the data available at the time of this study. Although not a gold standard, the K10 was a clear choice as a reference.
tool for the GSS due to its use as a screening tool for non-specific, global distress, similar to the trans-diagnostic nature of the GSS. As discussed in the previous section, future studies which aim to validate the GSS should only use the GAIN or another clinical assessment as the gold standard.

Finally, it is possible that the tools of interest (the GSS and GSS-I) share very similar levels of measurement error with their reference tools (K10, PHQ-9, and GAD-7). When someone responds to an item on a screening tool there will be a slight chance that they record the incorrect answer or select an answer that does not accurately describe their situation. When put in this situation four times with the GSS, K10, PHQ-9, and GAD-7, it is likely that an individual introduced the same amount of error in each tool. This limitation could have been avoided if the reference tool was a clinical assessment, such as the GAIN, similar to previous studies investigating the validity of the GSS. By using a clinical assessment, the probability of measurement error is still present, but due to its administration by someone other than the participant a different level of measurement error may be introduced. An added strength of using a clinical assessment such as a gold standard is that interviewers are allowed to provide clarification of items to participants to aid their understanding of the item.\textsuperscript{44} Along with the previous limitation, future studies could address this shortcoming by using a clinical assessment as the gold standard.

4.7 Applications of Research

Although the main objective this study was to contribute to the literature on the validity of the GSS, the applications of this research helped uncover information regarding the population which accesses services from an integrated youth health service. Based on a sample of 201 youth, over half the sample self-endorsed anxiety or depression. These findings are consistent
with multiple studies which have demonstrated that the prevalence of anxiety and depression are not only high but also commonly comorbid among youth. This study also revealed some information about the comorbidity of mental illness diagnoses in this sample. Women reported the greatest amount of multiple diagnoses with 50.4%, but alarmingly it appears that almost all participants who identified as non-binary in this sample reported multiple diagnoses. This finding is consistent with previous research that suggests that individuals who identify as transgender are at a greater risk of developing a mental illness. In a study of the mental health and substance use patterns among individuals who identified as transgender, Clements-Nolle et al. (2001) found that between 55% and 62% of transgender participants had depression and 32% of the sample reported attempting suicide. Individuals who identify as transgender and nonbinary are an important and emerging demographic in research due to the inclusion of gender identities besides women and men in study questionnaires. This application of research and how it pertains to the current study will be discussed next.

One area of this study that is noteworthy is that inclusion of data from individuals who are nonbinary. It is common in research to stratify data by sex and have data presented in terms of male and female. For this study, individuals were allowed to identify as men, women, or self-identify their gender if it fell outside of the gender binary. For all individuals who identified outside of the gender binary, this study is an individual act of gender representation in research. By not excluding their data we were able to provide some information on how nonbinary identified participants were characterized within this population. Further research is required to understand trends and health behaviors of nonbinary individuals and to determine if the trends are consistent or different than men and women. As an integrated youth health service, Foundry is an excellent position to cater to the needs of the LGBTQ population by creating an accepting and friendly environment to access health services. On their website, Foundry makes it clear that
understanding gender and sexual orientation are important to mental wellbeing. Based on the findings of this study, future researchers should consider LGBTQ mental health a priority and continue to include diverse gender and sexuality data in their work.
Chapter 5: Conclusion

5.1 Overall Significance and Contributions of Research

The main objectives of this study were to report the results of a detailed literature search on the GAIN and GSS, contribute to the existing literature on the GSS by examining its sensitivity and specificity, and make a novel contribution to the literature by also examining the sensitivity and specificity of the GSS-I. These contributions will help to advance the body of knowledge on the GAIN and the GSS, and their suitability as assessments and screening tools in integrated youth health services. The GAIN family of instruments have existed for over 20 years and have been used as assessment and screening tools in research for over 15 years, yet no definitive summary of the GAIN or GSS literature exists to inform the use of the tool. This study serves to address that gap by capturing the existing literature on the tool in a structured format and categorizing the types of articles based on the way the tools were implemented. The examination of sensitivity and specificity of the GSS and GSS-I also provides a contribution to the literature pertaining to the psychometric properties of the tool. Although the results were limited, they help explain the importance of carefully selecting a screening tool and what tools can act as appropriate gold standards.

In conclusion, the results of this study indicated that the GSS and GSS-I are tools with acceptable sensitivity and will likely flag the majority of individuals in a sample as having a moderate disorder. If the GSS is used in a screening process where a follow-up clinical assessment is readily available, the low specificity of the tool should not pose major consequences. In order for this to be possible, the ideal context for using the GSS would be within an integrated youth health service, where such resources exist. Ideally, the use of a tool for regular screening should be grounded in methodologically sound research and positive support from multiple sources. Based on the review of the literature there is some endorsement
of the GSS, and early studies suggested that it was valid for use among adolescents. This study contrasts the existing literature and encourages further evidence on the validity of the GSS.

There are clear limitations from this study that may have impacted the results, and these should also be taken into consideration when interpreting the usefulness of the GSS. Future studies should continue to validate the GSS and the GSS-I, as well as the remaining sub-scales in youth populations accessing an integrated youth health service.
References


3. Adolescent Development | Cleveland Clinic [Internet]. [cited 2019 Jun 1]. Available from: https://my.clevelandclinic.org/health/articles/7060-adolescent-development


8. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages


33. Rush B, Castel S, Desmond R. Screening for concurrent substance use and mental health problems in youth. Centre for Addiction and Mental Health; 2009.


41. Schulberg HC, Burns BJ. Mental Disorders in Primary Care: Epidemiologic, Diagnostic, and Treatment Research Directions* [Internet]. [cited 2019 May 14]. Available from: https://www-sciencedirect-com.ezproxy.library.ubc.ca/science/article/pii/0163834388900928


72. Background [Internet]. Foundry. 2019 [cited 2019 Mar 5]. Available from:
https://foundrybc.ca/who-we-are/background/


74. Who Are We [Internet]. Foundry. 2019 [cited 2019 Mar 5]. Available from: https://foundrybc.ca/who-we-are/


88. Pearson Product-Moment Correlation - When you should run this test, the range of values the coefficient can take and how to measure strength of association. [Internet]. [cited 2019 Apr 25]. Available from: https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php


91. Tavakol M, Dennick R. Making sense of Cronbach’s alpha. Int J Med Educ [Internet].


96. ROC curves – what are they and how are they used? [Internet]. [cited 2019 Apr 25]. Available from: https://acutecaretesting.org/en/articles/roc-curves-what-are-they-and-how-are-they-used


## Appendix A

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Type of Study</th>
<th>Study Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tims</td>
<td>2002</td>
<td>Characteristics and problems of 600 adolescent cannabis abusers in outpatient treatment</td>
<td>Measurement</td>
<td>• This appears to be one of the first studies to have used the GAIN in a setting where it was a primary measurement tool.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• The study employed the GAIN at baseline to determine diagnoses for Substance Use Disorders (SUD), Conduct Disorder (CD), and Attention Deficit Hyperactivity Disorder (ADHD).</td>
</tr>
<tr>
<td>Dennis</td>
<td>2002</td>
<td>The Cannabis Youth Treatment (CYT) experiment: rationale, study design and analysis plans</td>
<td>Validity</td>
<td>• The study used the GAIN at baseline and all other follow-up assessments.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• The authors only assessed past year symptoms at intake.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• The article discussed monetizing the elements of the GAIN to perform an economic analysis on the benefit/cost ratios of adolescents for each site.</td>
</tr>
<tr>
<td>Godley</td>
<td>2002</td>
<td>Preliminary outcomes from the assertive continuing care experiment for adolescents discharged from residential treatment</td>
<td>Validity</td>
<td>• Complete article unavailable.</td>
</tr>
<tr>
<td>Dennis</td>
<td>2003</td>
<td>An experimental evaluation of recovery management checkups (RMC) for people with chronic substance use disorders</td>
<td>Validity</td>
<td>• Complete article unavailable.</td>
</tr>
<tr>
<td>Titus</td>
<td>2003</td>
<td>Gender Differences in Victimization Severity and Outcomes Among Adolescents Treated for Substance Abuse</td>
<td>Gold Standard</td>
<td>• The GAIN-I and GAIN-M90 were used to collect outcome measures.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Study Title</td>
<td>Section</td>
<td>Notes</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Perry</td>
<td>2004</td>
<td>Adolescent and Young Adult Heroin and Non-Heroin Users: A Quantitative and Qualitative Study of Experiences in a Therapeutic Community</td>
<td>Measurement</td>
<td>- The GVI (General Victimization Index) is a 15-item tool that is part of the GAIN-I and was the item of interest in the validation analysis.</td>
</tr>
</tbody>
</table>
| Dennis     | 2004 | Cross-validation of the alcohol and cannabis use measures in the GAIN and TLFB among adolescents in substance abuse treatment. | Validity   | - The GAIN was used to collect baseline and intermittent outcome measures.  
- Results from this study indicated that heroin users had greater polysubstance use and lower self-efficacy following intervention. |
| Womack     | 2004 | Improving Treatment Services for Substance Abusers with Comorbid Depression | Measurement | - One of the most significant aspects of this study was how the GAIN was incorporated in conjunction with the DSM. |
First, participants were screened for depression using the GAIN. Then, those who met the criteria with the GAIN were subsequently screened with criteria from the DSM. Finally, those who were flagged by both tools were invited to participate in the intervention.

<table>
<thead>
<tr>
<th>Godley 2005</th>
<th>The stability and impact of environmental factors on substance use and problems after adolescent outpatient treatment for cannabis abuse or dependence.</th>
<th>Measurement</th>
<th>Used a variety of GAIN subscales to measure outcome variables of social risk, social support, substance use, recovery, environment risk, and substance related problems.</th>
</tr>
</thead>
</table>
| Lennox 2006 | Combining psychometric and biometric measures of substance use | Measurement | This study used biometric measures along with self-reported measures to build structural equation models that would demonstrate all of these individual factors are driven by substance use.  
  The main findings indicated that the GAIN Substance Frequency Scale performed as well or better than any other individual measures, and they also commented that the biometric measures were not as sensitive to the severity of other problems measured. |
<p>| Dennis 2006 | Development and Validation of the GSS for internalizing, externalizing, substance use disorders, and crime and violence problems among adolescents and adults. | Validity | This study was the first to validate the GSS and remains to be one of the most widely cited publications for referencing the GSS. |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Section</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stevens</td>
<td>2007</td>
<td>Effects of a behavioural sleep medicine intervention on trauma symptoms in adolescents recently treated for substance abuse</td>
<td>Measurement</td>
<td>The GAIN was administered to measure Post-Traumatic Stress Disorders (PTSD) symptoms</td>
</tr>
</tbody>
</table>
| Riley      | 2007 | Relative Precision, Efficiency and Construct Validity of Different Starting and Stopping Rules for a Computerized Adaptive Test: The GAIN Substance Problem Scale | Validity      | * The Computerized Adaptive Test (CAT) attempts to derive an algorithm which can determine the most efficient manner in which a screening test is delivered.  
  * Results showed a reduction in 13 to 66% of the 16 item SPS.  
  * Precision was lower overall, but changes were not significant and construct validity was not changed significant while improving efficiency.  
  * The authors suggest that this method could be applied to the full GAIN to determine a more efficient method of delivering the assessment because it takes 2 hours to administer currently. |
| Sacks      | 2007 | CJDATS Co-Occurring Disorders Screening Instrument For Mental Disorders | Gold Standard | * Appears to be the first study to use the GSS as a gold standard in the validation of another tool.  
  * The GSS was among two of the best performing instruments in this study, but all differences in performance were not statistically significant, with the tool of interest for this study performing as well if not better than the reference tools. |
<p>| Subramaniam| 2007 | Baseline Depressive Symptoms Predict Poor Substance Use Outcome       | Measurement   | * Used six factors from the GAIN to derive predictive risk factors for substance use.                                                      |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Title</th>
<th>Publication Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Following Adolescent Residential Treatment</td>
<td></td>
<td></td>
<td></td>
<td>- The GAIN was used to measure depressive symptoms and provide DSM diagnoses.</td>
</tr>
<tr>
<td>Dennis</td>
<td>2008</td>
<td>Global appraisal of individual needs—short screener (GAIN-SS): Administration and scoring manual version 2.0.3</td>
<td>Administration Manual (Other)</td>
<td>- This administration manual for the GSS outlines how to administer and score the tool, as well as some general guidelines for cut points across different age groups for the overall and sub-screening tools.</td>
</tr>
<tr>
<td>Gotham</td>
<td>2008</td>
<td>An Implementation Story: Moving the GAIN from Pilot Project to Statewide Use</td>
<td>Implementation Study (Other)</td>
<td>- An important article which details the wide-scale implementation of the GAIN.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- There were six stages of implementation: Exploration and Adoption, Program Installation, Initial Implementation, Full Operation, Innovation, and Sustainability.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- The basis of implementing a state-wide tool relies on commitment to the vision at all stages and the support of staff, treatment agencies, and state level support for decisions on funding and personnel resources.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- This model may be applied to future Canadian provinces wish to implement the GAIN.</td>
</tr>
<tr>
<td>Friedmann</td>
<td>2008</td>
<td>Violent and Disruptive Behavior among Drug-Involved Prisoners: Relationship with Psychiatric Symptoms</td>
<td>Measurement</td>
<td>- Used the GSS (excluding the substance use scale) to measure internalizing, externalizing, and crime and violence problems.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>Validity</td>
<td>Measurement</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Titus</td>
<td>2008</td>
<td>Development and Validation of Short Versions of the Internal Mental Distress and Behavior Complexity Scales in the Global Appraisal of Individual Needs (GAIN)</td>
<td>- Used a cut point of 5 for the GSS to screen for “severe mental health disorders”.</td>
<td>- Reported good sensitivity and specificity for diagnostic purposes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- The study used parallel tests of construct validity, as opposed to a more traditional validation approach using a gold standard.</td>
<td>- They compared short and long forms of comorbidity scales and subscales to demonstrate high correlation between the scales and parallel correlations with external criteria.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- The authors indicated the short scales have similar convergent and discriminate construct validity.</td>
<td>- The authors highlighted that individuals may still benefit from a full assessment, and it may be better to catch all potential cases then rule out negatives later.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Commented on how the use of a low cut point on the GAIN could lead to over identification.</td>
<td></td>
</tr>
<tr>
<td>Chan</td>
<td>2008</td>
<td>Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment</td>
<td></td>
<td>- The study demonstrates the importance of studying disorder comorbidity and how the GAIN is applied in this setting (presenting to substance abuse treatment).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- The main study findings are that youth had the highest co-occurring mental health problems, and that people with co-occurring mental health problems</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>Section</td>
<td>Details</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Titus      | 2008 | Characteristics of Youths With Hearing Loss Admitted to Substance Abuse Treatment | Measurement | - This study characterized a sample of individuals with hearing loss presenting to substance use treatment with the GAIN-I.  
- Findings indicated that individuals with hearing loss were generally a more vulnerable group reporting higher levels of distress. |
| Ramchand   | 2009 | Seven-Year Life Outcomes of Adolescent Offenders in Los Angeles       | Measurement | - Used the GAIN to interview participants, providing measurements for criminal involvement, substance use, and various internalizing and externalizing disorders.  
- Many participants died during the course of the study, with the authors recommending greater focus on rehabilitation programs. |
| McDonell   | 2009 | Global Appraisal of Individual Needs Short Screener (GSS): Psychometric Properties and Performance as a Screening Measure in Adolescents | Validity | - The second major study validating the GSS study examined 95 adolescents and reported the psychometric properties of the tool.  
- ROC Curves revealed sensitivity of 88% and specificity of 89%, where the gold standard used was an interview involving the full GAIN.  
- The authors determined an optimal cut point of 3 on the total disorder screener, other cut points for low, moderate, and high severity were consistent with the Dennis study. |
<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Study Description</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scott</td>
<td>2009</td>
<td>Results from two randomized clinical trials evaluating the impact of quarterly recovery management checkups with adult chronic substance users</td>
<td>- The GAIN was the primary tool used to measure outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Recovery Management Checkup (RMC) is a practice in which patients who undergo regular monitoring, screening, and intervention will experience earlier detection of relapse and better long-term outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- RMC was determined to be feasible and effective for adults who have substance dependence.</td>
</tr>
<tr>
<td>Conrad</td>
<td>2010</td>
<td>Screening for atypical suicide risk with person fit statistics among people presenting to alcohol and other drug treatment</td>
<td>- Used the GAIN to develop person fit statistics. Generally found that those who endorse at least one suicide symptom are also more likely to have some form psychiatric disorder.</td>
</tr>
<tr>
<td>Truman</td>
<td>2011</td>
<td></td>
<td>- Used the GSS in a cross-sectional study to rule out who had a behavioural disorder and who does not among domestic versus expatriate workers.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Participants were categorized using cut points across the low, moderate, and high scores. Expatriate workers had a greater risk of mental health and substance use disorders.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>Section</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
</tbody>
</table>
| Jessup     | 2011 | Validity and Reliability of the COJAC Screening Tool for Co-occurring Disorders | **Gold Standard**     | • The GSS was used as a reference tool for the validation of the COJAC. The authors make note that because the GSS has time sensitive items, 6 CST items were modified to include “in the past month.”  
• Highest Cronbach alpha scores were reported with the GSS. Some CST items were poorly correlated with the GSS and should be replaced. |
| Tarantino  | 2011 | Use Severity Mediates The Association Between Daily Marijuana Use And Psychiatric/Behavioral Risk In Emerging Adult Ed Patients: | **Measurement**       | • Complete article unavailable.                                                                                                                                                                        |
| Coleman-Cowger | 2012 | Mental Health Treatment Need Among Pregnant and Postpartum Women/ Girls Entering Substance Abuse Treatment | **Measurement**       | • GAIN administered upon entry into substance abuse treatment. No disparities found across age, but there were disparities in receiving treatment across race. Greater rates of mental illness in pregnant and post-partum women also reported. |
| Garnick    | 2012 | The Washington circle engagement performance measures’ association with adolescent treatment outcomes | **Measurement**       | • Outpatient treatment for SUD is the setting for this study and the GAIN was administered at intake and follow-up.  
• Engagement was the primary variable of interest for this study, and it was revealed that greater engagement with outpatient treatment facilities resulted in lower substance use.  
• The authors highlighted the use of the GAIN as a strength of their study. |
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jessup</td>
<td>2012</td>
<td>Unmet Mental Health and Substance Abuse Treatment Needs of Sexual Minority Elders</td>
<td>• Used an older version of the GSS which had fewer items than the current version. • Sexual minorities were bisexual, and reported greater levels of depression, anxiety, and suicidality.</td>
</tr>
<tr>
<td>Jessup</td>
<td>2012</td>
<td>Smoking and behavioral health of women</td>
<td>• The COJAC Screening Tool was used alongside the GSS in this study and was moderately correlated with the GSS (0.75) • The main findings are that smoking in women was associated with higher rates of mental illness and substance use problems. • The authors conclude that screening is important for women’s health.</td>
</tr>
<tr>
<td>McManus</td>
<td>2012</td>
<td>The impact of loneliness on sexual risk behaviors among African American women</td>
<td>• Dissertation which used the GSS as a screening tool to measure outcomes.</td>
</tr>
<tr>
<td>Claro</td>
<td>2012</td>
<td>[Translation and cultural adaptation of the Global Appraisal of Individual Needs - Initial]</td>
<td>• Prior to this, the GAIN was only administered in English and Spanish, so this appears to be the first study to translate the tool into another language. • Content validity was determined to be 0.91. The authors recommended actual tests of validity and reliability on the translated version.</td>
</tr>
<tr>
<td>Titus</td>
<td>2012</td>
<td>Impact of a training and certification program on the quality of interviewer-collected self-report assessment data</td>
<td>• Focuses on the certification process to administer the GAIN-I. • They found that individuals who had performed at least 15 administrations of the GAIN-I performed with higher...</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Study Description</td>
<td>Source</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Remy</td>
<td>2013</td>
<td>Correlates of unprotected sex in a sample of young club drug users</td>
<td>Measurement</td>
</tr>
<tr>
<td>Branson</td>
<td>2013</td>
<td>Text message reminders to improve outpatient therapy attendance among adolescents: A pilot study</td>
<td>Measurement</td>
</tr>
<tr>
<td>Felleman</td>
<td>2013</td>
<td>Behavioral health services influence medical treatment utilization among primary care patients with comorbid substance use and depression</td>
<td>Measurement</td>
</tr>
</tbody>
</table>

- fidelity than those who had less experience.
- The GAFI is the tool that was used to determine fidelity and increased with training.
- An updated version of the administration manual for the GSS with additional items, bringing the total number of items up to 23.
- The GAIN was the main survey tool and used to measure outcomes in a sample of elderly patients.
- Authors concluded that the elderly most likely misuse their own prescription medication to manage their pain.
- Data collection on drug use was completed using an abbreviated version of the GAIN. Authors found that unprotected sex was associated with alcohol and drug use.
- The GSS was administered at intake. This study found that text message reminders were affective for improving attendance to an outpatient substance use therapy program.
- Substance use was screened using the GSS. Depression increased the relationship between substance use and primary care service usage, and also behavioural health services were
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Study Description</th>
<th>Measurement</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarantino</td>
<td>2013</td>
<td>Family support mediates the association between substance use severity and suicidal ideation in early adult emergency department patients</td>
<td>Only one item was used from the GSS to assess suicidal ideation. The effect of cocaine use severity mediated by family support on suicidal ideation was positive and indirect.</td>
<td></td>
</tr>
<tr>
<td>Coleman-Cowger</td>
<td>2013</td>
<td>Comparison of the Addiction Severity Index (ASI) and the Global Appraisal of Individual Needs (GAIN) in predicting the effectiveness of drug treatment programs for pregnant and postpartum women</td>
<td>Findings from this study highlighted that the GAIN was moderately correlated with the Addiction Severity Index (ASI), has equal or greater coefficient alpha values than the ASI, and includes scales such as the HIV scale which are relevant for this population. Generally, the GAIN had coefficient alphas greater than 0.7 with the exception of one scale, indicating good internal consistency. Authors recommended that the ASI be validated with the shorter version of the GAIN as it may be more comparable.</td>
<td></td>
</tr>
<tr>
<td>Garner</td>
<td>2013</td>
<td>The GAIN short screener (GSS) as a predictor of future arrest or incarceration among youth presenting to substance use disorder (SUD) treatment</td>
<td>Found the GSS, crime and violence, and substance use screeners to be the best predicates of future incarceration out of the four sub-screeners in the tool. The GSS was also capable of sorting participants into three categories of risk: low, moderate, and high). The AUC value for the ROC curve was 0.61, indicting good predictive validity.</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>Section</td>
<td>Details</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Teodoro</td>
<td>2014</td>
<td>From Problem Focused To Holistic: Mental Health And Substance Abuse Screening At An Std Clinic</td>
<td>Measurement</td>
<td>Ultimately, the GSS performs as well as other tools in this domain at a fraction of the time.</td>
</tr>
<tr>
<td>Stucky</td>
<td>2014</td>
<td>A psychometric assessment of the GAIN Individual Severity Scale (GAIN-GISS) and Short Screeners (GAIN-SS) among adolescents in outpatient treatment programs</td>
<td>Validity</td>
<td>Those with higher rates of mental health concerns are more likely to have STIs. Participants were screened with the GSS, and those with high scores based on severity, quantity, and timing were linked to the appropriate resources.</td>
</tr>
<tr>
<td>McGarvey</td>
<td>2014</td>
<td>Effectiveness of A-CRA/ACC in Treating Adolescents with Cannabis-Use Disorders</td>
<td>Measurement</td>
<td>The third major study to validate the GSS investigated the psychometric properties of the tool through exploratory and item factory analysis. Three of the four subscales of the GSS were fit closely with unidimensional models, indicating that individual items measured individual constructs. Overall, the GSS was described as having more severe reductions in precision than favourable relative to its parent scales. The authors suggest that total disorder screener scores not be calculated often as their utility may not be as reliable as individual screener scores.</td>
</tr>
<tr>
<td>McGarvey</td>
<td>2014</td>
<td>Effectiveness of A-CRA/ACC in Treating Adolescents with Cannabis-Use Disorders</td>
<td>Measurement</td>
<td>The GAIN was administered at intake and follow-up to collect outcome measures. The intervention used in the study improved substance use rates and decreased time spent out of school.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Title</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
</tbody>
</table>
| Ratterman   | 2014 | Measuring the impact of substance abuse on student academic achievement and academic growth | - The GSS was administered to determine if participants were effectively moving towards becoming sober.  
- The authors acknowledge the detrimental impact of substance use on student’s school trajectory and support recovery schools for youth.                                                                                                                                                                                                                                                                                                                                                       |
| Schuler     | 2014 | Effectiveness of Treatment for Adolescent Substance Use: Is Biological Drug Testing Sufficient? | - All participants were interviewed with the GAIN at intake and at follow-up assessments. The GAIN also helped identify what services the participants were utilizing to determine treatment groups.  
- Authors comment on the subjectivity of self-report substance use data in the GAIN potentially introducing bias into the study, and that limitations in the crime and violence scales may not fully measure criminal involvement of participants.                                                                                                                                                                                                                                                                 |
| Chan        | 2014 | Referral for substance abuse treatment and depression improvement among patients with co-occurring disorders seeking behavioral health services in primary care | - The GSS was used to measure substance use severity.  
- The authors highlighted that the later a participant was admitted to treatment was associated with worsened depressive outcomes. Thus, early intervention is essential for co-occurring disorders.                                                                                                                                                                                                                                              |
<p>| Cerimele    | 2014 | Bipolar Disorder in Primary Care: Clinical Characteristics of 740 Primary Care Patients With Bipolar Disorder | - The GSS was used to measure substance use among patients, and a cut point of 3 indicated the high likelihood of a substance use disorder. The |</p>
<table>
<thead>
<tr>
<th>Authors/Year</th>
<th>Title</th>
<th>Measurement</th>
<th>Gold Standard</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balyakina 2014</td>
<td>Risk of Future Offense Among Probationers with Co-occurring Substance Use and Mental Health Disorders</td>
<td>Measurement</td>
<td>- The GSS was administered to screen for co-occurring disorders. Over one-quarter of the study sample screened positively on the internalizing scale when a cut-off score of 3 was used. The second greatest proportion of positive screens was on the substance use scale, with 23.3% of participants screening positively.</td>
<td></td>
</tr>
<tr>
<td>Veldhuizen 2014</td>
<td>“Do You Think You Have Mental Health Problems?”: Advantages and Disadvantages of a Single Screening Question for Mental Disorder in Substance Use Disorder Treatment</td>
<td>Gold Standard</td>
<td>- A single-item screening tool was validated among various screening tools, including the 20-item GSS. - The authors conclude that the single item tool poses an interesting area for future research, as it was able to perform moderately as well as the longer screening tools. - The authors also discuss that when resources are available full assessments are preferable to screening.</td>
<td></td>
</tr>
<tr>
<td>Patchell 2015</td>
<td>The Effect Of A Culturally Tailored Substance Abuse Prevention Intervention With Plains Indian Adolescents</td>
<td>Measurement</td>
<td>- Used the GAIN-Q to determine substance use outcomes. - The culturally-tailored intervention was shown to be effective among indigenous participants with substance use issues.</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Study Description</td>
<td>Measurement/Validity</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>Burgess</td>
<td>2015</td>
<td>A substance abuse intervention program at a large Russian manufacturing worksite</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• A Russian version of the GSS was used to collect the primary behavioural health outcomes for this study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This study used a workplace-based alcohol intervention and the GSS was used to determine the effects of this intervention.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aubry</td>
<td>2015</td>
<td>One-year outcomes of a randomized controlled trial of Housing First with ACT in five Canadian cities</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Used the GSS to measure substance use severity, where participants who endorsed two or more symptoms on the scale were categorized as having problematic substance use.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claro</td>
<td>2015</td>
<td>Rasch model of the GAIN substance problem scale among inpatient and outpatient clients in the city of Sao Paulo, Brazil</td>
<td>Validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This study validated the substance use scale of the Portuguese version of the GSS using a Rasch model.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The outcomes of this study indicated that the SPS was a valid scale but may have redundant items. The authors recommended including items with a wider range of severity.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claro</td>
<td>2015</td>
<td>Drug use, mental health and problems related to crime and violence: cross-sectional study</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The GSS was used to measure outcomes related to substance use and crime and violence.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The study highlighted the relationship between problematic alcohol use and mental health and crime and violence problems.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conrad</td>
<td>2015</td>
<td>Validation of the Full and Short-Form Self-Help Involvement Scale Against the Rasch Measurement Model</td>
<td>Validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Self-Help Improvement Scale (SHIS) is a subscale of the full GAIN, with a complete version in the GAIN and a newly developed short version of the SHIS.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Description</td>
<td>Source</td>
<td>Notes</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Richardson</td>
<td>2015</td>
<td>Implementation evaluation of a tablet-based waiting room survey of patient reported outcomes in an integrated youth health clinic</td>
<td>Measurement</td>
<td>Using a Rasch measurement model, the two versions of the two were validated against the model and with each other. Results indicated that both the long and short versions of the tool with adequate psychometric properties. The authors concluded that the long and short versions of the tool were unidimensional, reliable, and valid. Complete article unavailable.</td>
</tr>
<tr>
<td>Kaminer</td>
<td>2016</td>
<td>Screening, assessment, and treatment options for youths with a substance use disorder</td>
<td>Measurement</td>
<td>Complete article unavailable.</td>
</tr>
<tr>
<td>Chan</td>
<td>2016</td>
<td>Screening and follow-up monitoring for substance use in primary care: An exploration of rural–urban variations</td>
<td>Measurement</td>
<td>The GSS substance use scale was used to screen and monitor changes in substance use to document differences between rural and urban areas. Follow-up showed decreased rates of substance use. Authors reported that there is a disproportionate amount of substance misuse in rural communities relative to urban. Complete article unavailable.</td>
</tr>
<tr>
<td>Goodwin</td>
<td>2016</td>
<td>Substance abuse assessment</td>
<td>Book (Other)</td>
<td>Complete book unavailable.</td>
</tr>
<tr>
<td>Leslie</td>
<td>2016</td>
<td>Long-term health and psychosocial status of youth who received substance abuse treatment in adolescence</td>
<td>Measurement</td>
<td>The GSS was used to measure outcomes related to substance use and mental health, with items assessing suicidal ideation being removed. The authors found that young adults in the same reported problematic</td>
</tr>
</tbody>
</table>
| Henderson | 2016 | An independent replication of the Adolescent-Community Reinforcement Approach with justice-involved youth | **Measurement** | • Eligibility to enroll in the study was determined by endorsing alcohol use on the GSS (among other criteria).  
• The GAIN was used to determine services received, and client demographics and characteristics.  
• Following the intervention, youth participants demonstrated reduced substance use problems and the authors supported the clinical approach to the intervention. |
| Eisenberg | 2016 | Emotional health of lesbian, gay, bisexual and questioning bullies: Does it differ from straight bullies? | **Measurement** | • Adapted five items from the GSS to measure internal mental distress. |
| Carney | 2016 | Reliability of the GAIN-SS, CRAFTT and PESQ screening instruments for substance use among South African adolescents | **Validity** | • This was a very brief study on the reliability of the GSS, and it compared the GSS to two other brief screening tools. The study revealed the GSS to have adequate internal consistency and moderate levels of agreement.  
• The study did not use as full clinical assessment as a gold standard. The reliability was determined from test and retest measures. |
| Claro | 2016 | Internal Consistency and Convergent Validity of the Portuguese Versions of the Global Appraisal of Individual Needs-Initial and Short Screener: | **Validity** | • Internal consistency was evaluated using Cronbach’s alpha (0.88), and convergent validity was evaluated using Pearson’s correlation (0.80). |
| Cappelli 2016 | Transitioning youth into adult mental health and addiction services: An outcomes evaluation of the Youth Transition Project. | Measurement | • Authors say they used a 27-item version of the GSS, but this is most likely incorrect. In the study the GSS was used to measure outcomes.  
• 100% of youth in the sample were identified as having a serious psychiatric disorder. |
| Conway 2017 | Co-occurrence of tobacco product use, substance use, and mental health problems among adults: Findings from Wave 1 (2013–2014) of the Population Assessment of Tobacco and Health (PATH) Study | Measurement | • Used the GSS, modified for the PATH study. Additionally, the authors report the strengths of their study being that they used the GSS.  
• They also indicated that due to the omission on items measuring tobacco use the results may be limited as they cannot assess these associations.  
• Findings from this study indicated that females who smoke are at greater risk of developing substance use problems. |
<p>| Dell 2017 | Test Review: Global Appraisal of Individual Needs-Short Screener (GAIN-SS) | Review (Other) | • This review of the GSS is a brief report summarizing the history of the GSS, its properties, and the strengths and limitations of its use. |
| Ludmer 2017 | Accounting for the impact of parent internalizing symptoms on parent | Measurement | • This study used the GSS to assess internalizing symptoms. |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Title</th>
<th>Section</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tucker</td>
<td>2017</td>
<td>A group-based motivational interviewing brief intervention to reduce substance use and sexual risk behavior among homeless young adults</td>
<td>Measurement</td>
<td>- The results indicated that parents who endorse serious internalizing problems display conflicts with their ability to be a parent.</td>
</tr>
<tr>
<td>Kucharska</td>
<td>2017</td>
<td>Sex differences in the appraisal of traumatic events and psychopathology</td>
<td>Measurement</td>
<td>- The GSS was used to measure alcohol and marijuana usage.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Participants were sorted into lower or higher severity groups depending on their GSS scores under or greater than 3 respectively.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- The findings of this study indicated that the intervention can have positive effects on helping homeless adolescents make better choices surrounding alcohol and condom use.</td>
</tr>
<tr>
<td>Sloas</td>
<td>2017</td>
<td>Is Treatment Readiness Associated With Substance Use Treatment Engagement? An Exploratory Study</td>
<td>Measurement</td>
<td>- A modified version of the GAIN was used to measure mental disorder symptoms. The modified version was shorter than the original GAIN and was made with items from the GSS.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Following this study, it was revealed that women evaluated traumatic events more negatively than men.</td>
</tr>
<tr>
<td>Smith</td>
<td>2017</td>
<td>Sensitivity and specificity of the gain short-screener for predicting substance</td>
<td>Validity</td>
<td>- This study used data was collected using the GAIN-I on adults between the ages of 18 and 25.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Findings of the study do not support treatment readiness is related to treatment engagement.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- This can be seen as a follow-up study to an earlier study which highlighted the strengths of the substance use scale.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Study Description</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Cherner</td>
<td>2017</td>
<td>Housing first for adults with problematic substance use</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td>Csiernik</td>
<td>2017</td>
<td>Substance use of homeless and precariously housed youth in a Canadian context</td>
<td>Measurement</td>
<td></td>
</tr>
<tr>
<td>Henderson</td>
<td>2017</td>
<td>Screening for substance use and mental health problems in a cross-sectoral sample of Canadian youth</td>
<td>Measurement</td>
<td></td>
</tr>
</tbody>
</table>

- The gold standard appears to be the full criteria for a substance use disorder from the DSM IV.
- One of the limitations says the GSS substance use scale is drawn from the full version of the GAIN and recommends that future studies actually validate the GSS and its subscales with other screening tools.
- Only used substance disorder screener. Used the scale to determine admission criteria to the study.
- The substance use scale of the GSS was used to screen youth, and additional items on substance use age of initiation, money spent, and problems associated with substance use were also asked.
- Findings from this study indicated that precariously housed youth begin substance use at a lower age relative to stably house youth.
- Following the intervention there were observed decreases in substance use for both sexes.
- The GSS was used to measure outcomes in this study. The majority of the youth who presented to treatment facilities endorsed significant problems on the GSS that would likely meet the criteria for a diagnosis in a full clinical interview.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measurement</th>
</tr>
</thead>
</table>
| Remy   | 2017 | Anxiety and depression symptoms in Brazilian sexual minority ecstasy and LSD users | • This study used an abbreviated version of the GAIN-I which was culturally appropriate and developed to be used in street locations.  
• This study revealed greater rates of depression among sexual minorities relative to heterosexuals. |
| Smith  | 2017 | Housing instability and concurrent substance use and mental health concerns: An examination of Canadian youth | • The GSS was chosen for this study due to the fact that it screens over four domains that would help identify information about the precariously housed population.  
• The GSS was able to screen precariously housed youth outside of Mental Health and Addictions sectors, although the majority of youth were presenting to non-clinical sectors.  
• These studies highlight that issues related to service access and fragmentation does exist among these communities. |
| Coleman| 2017 | Describing the relationship between psychosocial distress and extent of criminal justice system involvement for men and women in a sober living environment | • The GSS was used to measure psychosocial distress and to study its interaction with gender differences and length of stay in a treatment facility.  
• The results of the study indicated that there were significant differences between gender. |
<p>| Gonzales| 2017 | Examining treatment outcomes: Differences between primary and non-primary marijuana-using youths transitioning into recovery | • Participants in a substance use treatment facility were screened with the GSS. The purpose was to study differences between those in treatment |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measurement</th>
</tr>
</thead>
</table>
| Grant  | 2017 | Developing adaptive interventions for adolescent substance use treatment settings: protocol of an observational, mixed-methods project | for marijuana use versus other substances.  
• Overall, youth who were in recovery for substances other than marijuana were more likely to report ceasing substance use. |
| Holliday | 2017 | Gender differences in the association between conduct disorder and risky sexual behavior | Using the full GAIN, this study collected outcome measures longitudinally which the authors believe helps to develop adaptive interventions.  
• The authors conclude that the research will help in developing multi-dimensional substance use treatment for adolescents. |
| Coleman | 2018 | Differences in Psychosocial Distress by Gender and Length of Residency in Criminal Justice System Involved Men and Women in a Sober Living Environment | This study only used the conduct disorder scale from the full GAIN, which was then constructed into a multivariate model to examine the relationship between GAIN scores and risky sexual behaviour. |
| Bell | 2018 | Predicting length of treatment involvement of substance using perinatal women in an integrated inpatient healthcare program | This appears to be one of the first studies to use the GSS in a criminal justice system in Oxford.  
• The authors noted a strong negative relationship between the length of stay in a criminal institution with all three subscales: internalizing, externalizing, and substance use problems.  
• This is a dissertation which used the GSS to measure mental distress among perinatal women. |
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measurement</th>
</tr>
</thead>
</table>
| Bruffaerts  | 2018 | Mental health problems in college freshmen: Prevalence and academic functioning | • The GSS was used to measure outcomes for this study. In the limitations for this study the authors highlighted that the GSS might not have been the best tool to screen for antisocial personality.  
• The authors link the presence of mental health concerns to college freshman, and lower academic performance. |
| Wilson      | 2018 | Identifying Predictors of Substance Use and Recidivism Outcome Trajectories Among Drug Treatment Court Clients | • Collected data using the GAIN but stated in their limitations that several items should be adjusted to collect more accurate data. |
| Lennox      | 2018 | Safety and tolerability of sauna detoxification for the protracted withdrawal symptoms of substance abuse | • The GSS was used alongside the ASI to measure outcome variables. The authors used a cut point of 1 on the GSS for mild to moderate severity, and a cut point of 3 for high severity.  
• Following the intervention there were improved mental health outcomes as measured by the GSS. |
| Conway      | 2018 | Co-occurrence of tobacco product use, substance use, and mental health problems among youth: Findings from wave 1 (2013–2014) of the population assessment of tobacco and health (PATH) study | • Used the GSS to measure outcome variables.  
• Authors stated that the use of the GSS was a strength due to high sensitivity and specificity of the tool. |
| Barnes      | 2018 | Emotional health among youth experiencing family homelessness          | • Used internalizing disorder screener from the GSS to measure outcomes.  
• These adolescents have over twice the risk of experiencing past year severe emotional distress compared to |
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choi</td>
<td>2018</td>
<td>Marijuana use among adults: Initiation, return to use, and continued use versus quitting over a one-year follow-up period</td>
<td>This study created survey items that were built from the mental health and substance use items from the GSS.</td>
</tr>
<tr>
<td>Amirloo</td>
<td>2018</td>
<td>The relationship between psychopathology by abusing smoking, alcohol, and drugs in boy teenagers of Ahvaz city</td>
<td>Mean age of starting cigarette smoking was 6.16 years and the highest period of abuse incidence is between 17 and 22 years. Authors recommend more preventative work done during this age range.</td>
</tr>
</tbody>
</table>
# Load Data
The data we will be using for this study comes from the SPOR Dataset, collected by Skye Babic and her team. It had an initial sample size of about 350, it was reduced down to 204 to capture all the participants who responded to the GSS. It was then reduced to 201 after three participants were removed due to providing no answers at all or no answers to the GSS. This was deemed okay to do following an analysis of missing data completed by Chris Richardson on SPSS.

#1) Attach Data
attach(MANUSCRIPT_DATA)

#2) Generate Study Variables

# Generate GSSI Cut Points
GSSI_Cat <- cut(GSSI, breaks=c(0,1,7), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSI_bin <- cut(GSSI, breaks=c(0,1,7), right=FALSE, labels=c("0", "1"))
GSSI_Cat1 <- cut(GSSI, breaks=c(0,2,7), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSI_Cat2 <- cut(GSSI, breaks=c(0,3,7), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSI_Cat3 <- cut(GSSI, breaks=c(0,4,7), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSI_Cat4 <- cut(GSSI, breaks=c(0,7,8), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))

# Generate PHQ9 Cut Points
PHQ9_Cat <- cut(PHQtotal, breaks=c(0,5,28), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
PHQ9_bin <- cut(PHQtotal, breaks=c(0,5,28), right=FALSE, labels=c("0", "1"))

# Generate GAD7 Cut Points
GAD7_Cat <- cut(GAD7total, breaks=c(0,5,22), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GAD7_bin <- cut(GAD7total, breaks=c(0,5,22), right=FALSE, labels=c("0", "1"))

#3) Descriptive Statistics
# Age
mean(Age, na.rm = TRUE)
sd(Age, na.rm = TRUE)
hist(Age)
range(Age, na.rm=TRUE)
#Ignored missing variables

#Gender
gendertable <- table(MANUSCRIPT_DATA$Gender)
prop.table(gendertable)
table(Gender)
plot(Gender)
#Female, Male, Other, and Non-Response

#Education
#1=Some Highschool
#2=High School Diploma
#3=Some College/Technical School
#4=Some University
#5=College or Technical Degree
#6=Bachelor's Degree
#7=Master's Degree or Higher

MANUSCRIPT_DATA$Education <- factor(MANUSCRIPT_DATA$Education,
levels=c(1,2,3,4,5,6,7),
labels=c("Some High School", "High School Diploma", "Some College or Technical School",
"Some University", "College or Technical Degree", "Bachelor's Degree", "Master's Degree or Higher"))

table(MANUSCRIPT_DATA$Education)
educationtable <- table(MANUSCRIPT_DATA$Education)
prop.table(educationtable)
hist(Education)

#Ethnicity
frequency(Ethnicity)
ethnicitytable <- table(MANUSCRIPT_DATA$Ethnicity)
prop.table(ethnicitytable)
table(Ethnicity)

#Diagnoses
table(Diagnoses...choice.Anxiety.,Gender)
table(Diagnoses...choice.Bipolar.,Gender)
table(Diagnoses...choice.Depression.,Gender)
table(Diagnoses...choice.Other.,Gender)
table(Diagnoses...choice.Post.traumatic.stress.disorder., Gender)
table(Diagnoses...choice.Schizoaffective.disorder.,Gender)
table(Diagnoses...choice.Schizophrenia.,Gender)

#Substance Use
#1=No Substance Use
#2=Once a Month
#3=1-2 Times a Week
#4=3-5 Times a Week
#5=6-7 Times a Week

MANUSCRIPT_DATA$Substance.Use<-factor(MANUSCRIPT_DATA$Substance.Use, levels=c(1,2,3,4,5), labels=c("No Substance Use", "Once a Month", "1-2 Times a Week", "3-5 Times a Week", "6-7 Times a Week"))

table(Substance.Use,Substance.Use)
table(MANUSCRIPT_DATA$Substance.Use,Gender)

#By Gender
table(Education, Gender)
table(Ethnicity, Gender)
table(Age, Gender)

#2) Correlation Statistics
#Will need to create a new variable for the GSSI
GSSI<-c(6,6,6,3,3,4,5,5,6,6,6,5,5,4,5,6,6,5,6,6,3,6,6,4,4,5,0,6,6,6,4,3,5,3,1,6,4,6,4,5,5,6,6,6,5,6,5,6,5,6,6,6,6,4,6,5,2,2,0,5,4,4,4,5,4,6,3,5,5,5,2,5,6,6,4,6,5,0,5,6,6,4,4,0,5,4,5,3,5,5,0,5,6,6,0,5,3,5,6,5,0,6,2,6,3,4,6,3,6,5,3,4,5,5,6,4,6,5,5,6,2,5,6,5,5,2,6,4,6,5,3,2,5,4,6,4,0,0,4,4,6,4,5,2,4,5,0,6,6,3,6,5,4,6,5,6,5,6,6,4,0,5,6,7,4,5,3,4,3,5,4,6,6,3,4,5,5)

#Correlational Matrix
cor.test(GAINSStotaldisorderscreener, k10total, method="pearson")
cor.test(GSSI, PHQtotal, method="pearson")
cor.test(GSSI, GAD7total, method="pearson")
cor.test(GAINSStotaldisorderscreener,PHQtotal, method="pearson")
cor.test(GAINSStotaldisorderscreener, GAD7total, method="pearson")
cor.test(GAINSStotaldisorderscreener, GSSI, method="pearson")
cor.test(k10total, GAINSStotaldisorderscreener, method="pearson")
cor.test(k10total, PHQtotal, method="pearson")
cor.test(k10total, GAD7total, method="pearson")
cor.test(k10total, GSSI, method="pearson")
cor.test(PHQtotal, GAD7total, method="pearson")
cor.test(PHQtotal, GSSI, method="pearson")
cor.test(GAD7total, GSSI, method="pearson")

#All tests for significance came back as less than 0.01  
#Cronbach's Alpha
library(psych)
GSSmatrix <- matrix(c(GAINI1, GAINI2, GAINI3, GAINI4, GAINI5, GAINI6, GAINI7, GAINI8, GAINI9, GAINI10, GAINI11, GAINI12, GAINI13, GAINS14, GAINS15, GAINS16, GAINS17, GAINS18, GAINCV19, GAINCV20, GAINCV21, GAINCV22, GAINCV23), nrow = 201)
K10matrix <- matrix(c(k101, k102, k103, k104, k105, k106, k107, k108, k109, k1010), nrow=201)
GSSIImatrix <- matrix(c(GAINI1, GAINI2, GAINI3, GAINI4, GAINI5, GAINI6), nrow=201)
PHQ9matrix <- matrix(c(PHQ1, PHQ2, PHQ3, PHQ4, PHQ5, PHQ6, PHQ7, PHQ8, PHQ9), nrow=201)
GAD7matrix <- matrix(c(GAD1, GAD2, GAD3, GAD4, GAD5, GAD6, GAD7), nrow=201)

alpha(GSSmatrix)
alpha(K10matrix)
alpha(GSSIImatrix)
alpha(PHQ9matrix)
alpha(GAD7matrix)

#3) Validation

#GSS with K10
GSS_Categorical <- cut(GAINSStotaldisorderscreener, breaks=c(0, 1, 24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSS_bin <- cut(GAINSStotaldisorderscreener, breaks=c(0, 1, 24), right=FALSE, labels=c("0", "1"))

print(GSS_Categorical)[1:10]
GAINSStotaldisorderscreener[1:10]

K10_Categorical <- cut(k10total, breaks=c(0, 20, 51), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
K10_bin <- cut(k10total, breaks=c(0, 20, 51), right=FALSE, labels=c("0", "1"))

K10_Cat1 <- cut(k10total, breaks=c(0, 24, 51), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Categorical, K10_Categorical)

GSSCat1 <- cut(GAINSStotaldisorderscreener, breaks=c(0, 7, 24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSCat2 <- cut(GAINSStotaldisorderscreener, breaks=c(0, 8, 24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSCat3 <- cut(GAINSStotaldisorderscreener, breaks=c(0, 9, 24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSCat4 <- cut(GAINSStotaldisorderscreener, breaks=c(0, 23, 24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
GSSCat5 <- cut(GAINSStotaldisorderscreener, breaks=c(0,10,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))

#Cut Point 5#
GSS_Cat5 <- cut(GAINSStotaldisorderscreener, breaks=c(0,5,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Cat5, K10_Categorical)
149/(149+0)
1-(14/(14+38))

#Cut Point 6#
GSS_Cat6 <- cut(GAINSStotaldisorderscreener, breaks=c(0,6,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Cat6, K10_Categorical)
148/(148+1)
1-(16/(16+36))

#Cut Point 10#
GSS_Cat10 <- cut(GAINSStotaldisorderscreener, breaks=c(0,10,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Cat10, K10_Categorical)
116/(33+116)
1-(27/(25+27))

#Cut Point 15#
GSS_Cat15 <- cut(GAINSStotaldisorderscreener, breaks=c(0,15,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Cat15, K10_Categorical)
59/(59+90)
1-(38/(38+14))

#Cut Point 20#
GSS_Cat20 <- cut(GAINSStotaldisorderscreener, breaks=c(0,20,24), right=FALSE, labels=c("Symptoms Absent", "Symptoms Present"))
table(GSS_Cat20, K10_Categorical)
18/(18+131)
1-(46/(46+6))

#GSS with K10
table(GSS_Categorical, K10_Categorical)

#Sensitivity
149/(149+0)

#Specificity, FPR
5/(5+47)
1-(5/(5+47))

#PPV
149/(149+47)
#NPV
5/(5+0)

table(GSSCat1, K10_Categorical)
#Sensitivity
143/(143+6)
#Specificity, FPR
21/(21+31)
1-(21/(21+31))
#PPV
143/(143+31)
#NPV
21/(21+6)

table(GSSCat2, K10_Categorical)
#Sensitivity
135/(135+14)
#Specificity, FPR
22/(22+30)
1-(22/(22+30))
#PPV
135/(30+135)
#NPV
22/(22+14)

table(GSSCat3, K10_Categorical)
#Sensitivity
131/(131+18)
#Specificity, FPR
23/(23+29)
1-(23/(23+29))
#PPV
131/(131+29)
#NPV
23/(23+18)

table(GSSCat4, K10_Categorical)
#Sensitivity
6/(143+6)
#Specificity, FPR
48/(48+4)
1-(48/(48+4))

table (GSSCat5, K10_Categorical)
#Sensitivity
116/(116+33)
#Specificity, FPR
\[
\frac{27}{27+25} \\
1 - \frac{27}{27+25} \\
\text{#PPV}
\]

\[
\frac{116}{116+25} \\
\text{#NPV}
\]

\[
\frac{27}{27+33}
\]

#Generate Vectors
GSS_Sens <- c(1, 1, 0.9933, 0.9597, 0.9060, 0.8792, 0.7785, 0.394, 0.1208, 0.040, 0)
GSS_FPR <- c(1, 0.9038, 0.7308, 0.6923, 0.5962, 0.5769, 0.5577, 0.4808, 0.2692, 0.1154, 0.077, 0)

#ROC Curve
plot(type="b", GSS_FPR, GSS_Sens, xlab="False Positive Rate", ylab="Sensitivity")

#GSSI with PHQ9

table(GSSI_Cat, PHQ9_Cat)
#Sensitivity
\[
\frac{176}{176+4}
\]
#Specificity, FPR
\[
\frac{7}{7+14} \\
1 - \frac{7}{7+14} \\
\text{#PPV}
\]
\[
\frac{176}{176+14} \\
\text{#NPV}
\]
\[
\frac{7}{7+4}
\]

table(GSSI_Cat1, PHQ9_Cat)
#Sensitivity
\[
\frac{176}{176+4}
\]
#Specificity, FPR
\[
\frac{8}{8+13} \\
1 - \frac{8}{8+13} \\
\text{#PPV}
\]
\[
\frac{176}{176+13} \\
\text{#NPV}
\]
\[
\frac{8}{8+4}
\]

table(GSSI_Cat2, PHQ9_Cat)
#Sensitivity
\[
\frac{170}{170+10}
\]
#Specificity, FPR
\[
\frac{10}{10+11} \\
1 - \frac{10}{10+11} \\
\text{#PPV}
\]
\[
\frac{170}{170+11}
\]

122
#NPV
10/(10+10)

table(GSSI_Cat3, PHQ9_Cat)
#Sensitivity
156/(156+24)
#Specificity, FPR
14/(14+7)
1-(14/(14+7))
#PPV
156/(156+7)
#NPV
14/(14+24)

table(GSSI_Cat4, PHQ9_Cat)
#Sensitivity
0
#Specificity, FPR
21/(21+0)
1-(21/(21+0))

#Generate Vectors
GSSI_sens_PHQ9 <- c(1, 0.9778, 0.9778, 0.9444, 0.8667, 0)
GSSI_FPR_PHQ9 <- c(1, 0.6667, 0.6190, 0.5238, 0.3333, 0)

#ROC Curve
plot(type="b", GSSI_FPR_PHQ9, GSSI_sens_PHQ9, main="ROC Curve for Validation of the GSSI with the PHQ-9", xlab="False Positive Rate", ylab="Sensitivity")

#GSSI with Gad7
table(GSSI_Cat, GAD7_Cat)
#Sensitivity
172/(172+3)
#Specificity, FPR
8/(8+18)
1-(8/(8+18))
#PPV
172/(172+18)
#NPV
8/(8+3)

table(GSSI_Cat1, GAD7_Cat)
#Sensitivity
172/(172+3)
#Specificity, FPR
9/(9+17)
1-(9/(9+17))
# PPV
172/(172+17)

# NPV
9/(9+3)

table(GSSI_Cat2, GAD7_Cat)
# Sensitivity
165/(165+10)
# Specificity, FPR
10/(10+16)
1-(10/(10+16))
# PPV
165/(165+16)
# NPV
10/(10+10)

table(GSSI_Cat3, GAD7_Cat)
# Sensitivity
151/(151+24)
# Specificity, FPR
14/(14+12)
1-(14/(14+12))
# PPV
151/(151+12)
# NPV
14/(14+24)

table(GSSI_Cat4, GAD7_Cat)
# Sensitivity
1
# Specificity, FPR
1
0

GSSI_sens_GAD7 <- c(1, 0.9829, 0.9829, 0.9429, 0.8629, 0)
GSSI_FPR_GAD7 <- c(1, 0.6923, 0.6538, 0.6154, 0.4615, 0)

plot(type="b", GSSI_FPR_GAD7, GSSI_sens_GAD7, main="ROC Curve for Validation of the GSS-I with the GAD-7", xlab="False Positive Rate", ylab="Sensitivity")

# GSSI with K10
table(GSSI_Cat, K10_Categorical)
# Sensitivity
148/(148+1)
# FPR
1-(10/(10+42))
#PPV
148/(148+42)

#NPV
10/(10+1)

table(GSSI_Cat1, K10_Categorical)
#Sensitivity
148/(148+1)
#FPR
1-(11/(11+41))
#PPV
148/(148+41)
#NPV
11/(11+1)

table(GSSI_Cat2, K10_Categorical)
#Sensitivity
145/(145+4)
#FPR
1-(16/(16+36))
#PPV
145/(145+36)
#NPV
16/(16+4)

table(GSSI_Cat3, K10_Categorical)
#Sensitivity
136/(136+13)
#FPR
1-(25/(25+27))
#PPV
136/(136+27)
#NPV
25/(25+13)

table(GSSI_Cat4, K10_Categorical)
#Sensitivity
0/149
#FPR
1-(52/(52+0))

#Create Vectors
GSSI_Sens_K10 <- c(1, 0.9933, 0.9933, 0.9732, 0.9128, 0)
GSSI_FPR_K10 <- c(1, 0.8077, 0.7885, 0.6923, 0.5192, 0)

#ROC Curve
plot(GSSI_FPR_K10, GSSI_Sens_K10, type="b")
#AUC
#Load library 'MESS'
library(MESS)

#Generate AUC Variables
x<-GSS_FPR
y<-GSS_Sens
x1<-GSSI_FPR_K10
y1<-GSSI_Sens_K10
x2<-GSSI_FPR_PHQ9
y2<-GSSI_sens_PHQ9
x3<-GSSI_FPR_GAD7
y3<-GSSI_sens_GAD7

#GSS with K10
auc(x,y)

#GSSI with K10
auc(x1,y1)

#GSSI with PHQ9
auc(x2,y2)

#GSSI with GAD7
auc(x3,y3)
Appendix C

Figure 6 - Normal QQ-plot displaying that the data for the GSS is approximately normally distributed. This may also be used to demonstrate the data for the GSS-I is normally distributed as the GSS is the parent tool.

Figure 7 - Normal QQ-plot displaying that the data for the K10 is approximately normally distributed.
Figure 8 – Normal QQ-plot displaying that the data for the PHQ-9 is approximately normally distributed.

Figure 9 – Normal QQ-plot displaying that the data for the GAD-7 is approximately normally distributed.
Appendix D

Figure 10 – Relationship of GSS and K10 total scores, where a slight curve is present, but the overall shape is still monotonic. As a result, these scores can be considered linear.

Figure 11 – Relationship of GSS-I and K10 total scores where there is a clear linear relationship between the scores.
Figure 12 – Relationship of GSS-I and PHQ-9 total scores, where a slight curve is present but the overall shape is still monotonic. As a result, these scores can be considered linear.

Figure 13 – Relationship of GSS-I and GAD-7 total scores, where a slight curve is present but the overall shape is still monotonic. As a result, these scores can be considered linear.
Appendix E

GAIN Short Screener (GAIN-SS)
Version [GVER]: GAIN-SS ver. 3.0.1

What is your name?  a. ________ b. ________ c. ________
                      (First name)      (ML)        (Last name)

What is today’s date? (MM/DD/YYYY) ______/_____/______ / 20 ______

The following questions are about common psychological, behavioral, and personal problems. These problems are considered significant when you have them for two or more weeks, when they keep coming back, when they keep you from meeting your responsibilities, or when they make you feel like you can’t go on.

After each of the following questions, please tell us the last time, if ever, you had the problem by answering whether it was in the past month, 2 to 3 months ago, 4 to 12 months ago, 1 or more years ago, or never.

<table>
<thead>
<tr>
<th></th>
<th>Past month</th>
<th>2 to 3 months ago</th>
<th>4 to 12 months ago</th>
<th>1 year ago</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDScr 1</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EDScr 2  When was the last time that you did the following things two or more times?

<table>
<thead>
<tr>
<th></th>
<th>Past month</th>
<th>2 to 3 months ago</th>
<th>4 to 12 months ago</th>
<th>1 year ago</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SDScr 3  When was the last time that you...

<table>
<thead>
<tr>
<th></th>
<th>Past month</th>
<th>2 to 3 months ago</th>
<th>4 to 12 months ago</th>
<th>1 year ago</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td>3 2 1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 14 – Global Appraisal of Individual Needs (GSS) questionnaire with scoring guide for the disorder screener and the sub-screeners, GAIN-SS copyright © Chestnut Health Systems.
**K10 Test**

These questions concern how you have been feeling over the past 30 days. Tick a box below each question that best represents how you have been.

<table>
<thead>
<tr>
<th>1. During the last 30 days, about how often did you feel tired out for no good reason?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. During the last 30 days, about how often did you feel nervous?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. During the last 30 days, about how often did you feel so nervous that nothing could calm you down?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. During the last 30 days, about how often did you feel hopeless?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. During the last 30 days, about how often did you feel restless or fidgety?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. During the last 30 days, about how often did you feel restless you could not sit still?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. During the last 30 days, about how often did you feel depressed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. During the last 30 days, about how often did you feel that everything was an effort?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. During the last 30 days, about how often did you feel so sad that nothing could cheer you up?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. During the last 30 days, about how often did you feel worthless?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. None of the time</td>
</tr>
</tbody>
</table>
Figure 15 – Kessler Psychological Distress Scale (K10) questionnaire with scoring guide and recommended clinical cut points, developed by Kessler et al. (2002).
### GAD-7 Anxiety

Over the last 2 weeks, how often have you been bothered by the following problems?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Several days</th>
<th>More than half the days</th>
<th>Nearly every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeling nervous, anxious or on edge</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Not being able to stop or control worrying</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Worrying too much about different things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Trouble relaxing</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Being so restless that it is hard to sit still</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Becoming easily annoyed or irritable</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Feeling afraid as if something awful might happen</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Column totals: ___ + ___ + ___ + ___ = Total Score ___

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

<table>
<thead>
<tr>
<th>Not difficult at all</th>
<th>Somewhat difficult</th>
<th>Very difficult</th>
<th>Extremely difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues. For research information, contact Dr. Spitzer at rls8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc. Copyright © 1999 Pfizer Inc. All rights reserved. Reproduced with permission.
## PHQ-9 Depression

Over the **last 2 weeks**, how often have you been bothered by any of the following problems?

*(Use "✓" to indicate your answer)*

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>Several days</th>
<th>More than half the days</th>
<th>Nearly every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Little interest or pleasure in doing things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Feeling down, depressed, or hopeless</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. Trouble falling or staying asleep, or sleeping too much</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. Feeling tired or having little energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. Poor appetite or overeating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. Trouble concentrating on things, such as reading the newspaper or watching television</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. Thoughts that you would be better off dead or of hurting yourself in some way</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Column totals**  

**= Total Score**

---

From the Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PRIME-MD PHQ). The PHQ was developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues. For research information, contact Dr. Spitzer at rfs8@columbia.edu. PRIME-MD® is a trademark of Pfizer Inc. Copyright© 1999 Pfizer Inc. All rights reserved. Reproduced with permission.
Figure 16 – Patient Health Questionnaire 9 (PHQ-9) and Generalized Anxiety Disorder 7 (GAD-7) questionnaires with scoring guide and recommended clinical cut points, from Pfizer Inc. copyright © 1999 Pfizer Inc.