A MULTI-SITE EVALUATION OF CONTACT-BASED ANTI-STIGMA PROGRAMS FOR HIGH SCHOOL YOUTH

by

Michaela Koller

A thesis submitted to the Department of Public Health Sciences
In conformity with the requirements for the
degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(December 2018)

Copyright © Michaela Koller, 2018
Abstract

Background: Adolescents have become an important target for contact based anti-stigma programs. Research on mental health stigma reduction in adolescents has been hindered by lack of validated measures to evaluate program interventions. If we are to systematically reduce stigma and improve mental health, we must have the capacity to observe and measure stigma and to evaluate whether or not program interventions are producing the desired effect. The many scales used to measure attitudes towards people with mental illness that exist today are not adequate as they have been developed for the adult population outside of the intervention setting.

Objectives: The first objective was to develop two scales within a Utilization Focused Evaluation framework. One scale to measure stereotypic attributions toward people with a mental illness and another to measure social acceptance of those with a mental illness. The second objective was to use the newly developed scales to evaluate three different types of contact-based interventions that were delivered by one program.

Methods: Survey data were collected on three samples of adolescent students aged between 12 and 19 years old (total N=2118). Qualitative methods, including expert opinion and focus group discussion were used for item generation, item review and refinement, and in general to establish content validity. Quantitative methods, such as exploratory and confirmatory factor analysis, known group differences, Cronbach’s alpha and test-retest reliability were also used to assess validity and reliability. Using the standardized scales that were developed, 29 contact-based stigma reduction programs were evaluated with a non-randomized trial design to measure changes. For the second objective survey data were collected from 2146 adolescents who
received one of three contact-based interventions or a comparison situation. Data were analyzed using linear regression to evaluate the effects of the interventions.

Results: For the first objective, results showed that a one-factor solution was the best fit for the data for each scale. Following exploratory factor analysis, a confirmatory factor analysis was conducted on a different sample. Results highlighted some items with low loadings. Allowing for several correlated variances resulted in a good fitting model. These results in combination with the known group differences and the test – retest evaluation support preliminary reliability and validity of the two scales. When programs were evaluated, larger improvements were seen on the Stereotypic Attributions scale than on the Social Acceptance scale. For the second objective all three interventions types showed a positive change in social acceptance compared to a group that did not receive an intervention.

Conclusion: This study has established initial reliability and validity of two scales to evaluate youth attitudes towards and social acceptance of those with a mental illness in a wide range of samples of students from middle and high schools in Canada. We identified two unidimensional scales. The Stereotypic Attribution Scale contains 11 items measuring stereotypic attitudes towards people with a mental illness. The Social Acceptance Scale contains 11 items addressing social distance and social responsibility. This study has also demonstrated that contact-based education provided in class, as an assembly, or as the focus of a student summit can be promising practices for reducing stigma in middle and high school students.
Co-Authorship

This thesis presents the work of Michaela Koller in collaboration with her supervisor Dr. Heather Stuart, and her advisors Dr. Terry Krupa and Dr. Shu-Ping Chen. Michaela Koller coordinated the data collection and data entry, completed the statistical analyses, interpreted results, and wrote the thesis document with frequent edits and guidance from her committee.

Chapter 3 – Manuscript 1: Youth Attitudes Towards and Social Acceptance of the Mentally Ill: Development and Psychometric Analysis of Two Scales. This manuscript was formatted for submission to Open BMC Psychology. The need to develop these scales stemmed from desire to evaluate contact based mental health stigma reduction programs directed towards youth. The co-author on this manuscript is Heather Stuart. Interpretation of results and writing of the manuscript were done by Michaela Koller with supervision from Heather Stuart and feedback from Shu-Ping Chen.

Chapter 4 – Manuscript 2: A Utilization Focused Evaluation of Youth Contact-Based Stigma Reduction. This manuscript was formatted as a practice note for submission to the Canadian Journal of Program Evaluation. The co-author on this manuscript is Heather Stuart. Writing of the manuscript was done by Michaela Koller with supervision and feedback from Heather Stuart.

Chapter 5 Manuscript 3: Reducing Mental Health Stigma in Middle and High School Adolescents: A non-randomized trial. This manuscript was formatted for submission to Open BMC Psychology. The co-author on this manuscript is Heather Stuart. Interpretation of results and writing of the manuscript were done by Michaela Koller with supervision from Heather Stuart.
Acknowledgements

I would like to express my profound gratitude to Dr. Heather Stuart. I am extremely grateful for all your support. Your mentorship and faith in my abilities have been invaluable; without your encouragement I would never have even thought to start a PhD. I cannot thank you enough for your patience and guidance over the past years. Your thoughtful feedback and suggestions have been invaluable. Thank you to my advisors, Dr. Terry Krupa and Dr. Shu-Ping Chen, who have provided valuable feedback throughout the course of this project and have made this memorable experience. I would also like to thank those in the Department of Community Health and Epidemiology/Public Health Sciences for their support over the years. Further, I want to thank my friends and colleagues at CHSPR for their support. Thank you to friends and colleagues from Opening Minds, our programs partners, schools and students who participated in this project. Without them this project could not have been done. This work would not have been possible without financial support from the Mental Health Commission of Canada. I would also like to thank my family and extend a special thanks to my children for enduring my academic pursuits. I am grateful for their love, support and patience throughout this journey.
# Table of Contents

Abstract ..................................................................................................................................................... ii  
Co-Authorship........................................................................................................................................ iv  
Acknowledgements ................................................................................................................................... v  
Table of Contents ..................................................................................................................................... vi  
List of Tables ........................................................................................................................................... ix  
List of Figures .......................................................................................................................................... x  
List of Abbreviations ............................................................................................................................... xi  
Chapter 1 Introduction and Background ............................................................................................... 1  
  1.1 Purpose ........................................................................................................................................... 1  
  1.2 Objectives: ...................................................................................................................................... 2  
    1.2.1 Scale Development ................................................................................................................ 2  
    1.2.2 Outcome Evaluation ............................................................................................................... 3  
  1.3 Thesis Organization ...................................................................................................................... 5  
Chapter 2 – Background ............................................................................................................................ 7  
  2.1 Mental Illness Stigma .................................................................................................................... 7  
  2.2 Epidemiology of Stigma in Youth ................................................................................................. 11  
  2.3 Stigma Reduction ........................................................................................................................... 13  
  2.4 Social Contact Theory ................................................................................................................... 14  
  2.5 The nature of the interventions under study ............................................................................. 16  
  2.6 Review of Instrument Development ........................................................................................... 20  
    2.6.1 Item Generation and Scale Construction ............................................................................. 20  
    2.6.2 Validity and Reliability .......................................................................................................... 21  
      2.6.2.1 Validity ........................................................................................................................... 21  
      2.6.2.2 Reliability – Internal Consistency ................................................................................... 22  
      2.6.2.3 Reliability – Test-Retest ............................................................................................... 23  
      2.6.2.4 Validity - Confirmatory Factor Analysis ....................................................................... 24  
    2.6.3 Sample Size Considerations .................................................................................................. 24  
      2.6.3.1 Exploratory Factor Analysis ........................................................................................... 24  
      2.6.3.2 Confirmatory Factor Analysis ......................................................................................... 25  
  2.7 Program Evaluation ...................................................................................................................... 25  
    2.7.1 Research and Program Evaluation ........................................................................................ 25
List of Tables

Table 1 Sample Size .............................................................................................................................. 19
Table 2 Characteristics for the Exploratory, Confirmatory and the Test-Retest samples ................. 58
Table 3 Stereotypic Attribution factor loadings from exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and re-specified confirmatory factor analysis (CFA2) ........................................ 59
Table 4 Social Acceptance factor loading from exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and re-specified confirmatory factor analysis (CFA2) ........................................ 60
Table 5. Examples of Some of the Knowledge Translation Activates .................................................. 77
Table 6 Comparison Between Those Included and Excluded from Analyses Due to Inability to Match Baseline and Follow-up Surveys ........................................................................................................ 97
Table 7 Mean baseline scale scores .................................................................................................. 98
Table 8 Comparison between Intervention and the Comparison group ........................................... 99
Table 9 Mean baseline scale scores interventions vs comparison group ...................................... 100
Table 10 Intraclass Correlations ..................................................................................................... 100
Table 11 Bivariate linear relationships ............................................................................................ 102
Table 12 Final Model Based on the Linear Regression ................................................................. 104
List of Figures

Figure 1  Two External Stigma Models. .......................................................................................... 10
Figure 2  Program Logic Model.................................................................................................... 18
Figure 3  Program Logic Model.................................................................................................... 73
Figure 4  Data exclusion ............................................................................................................... 95
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
</tr>
<tr>
<td>CFI</td>
<td>Comparative Fit Index</td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass Correlation Coefficient</td>
</tr>
<tr>
<td>KMO</td>
<td>Kaiser-Meyer-Olkin</td>
</tr>
<tr>
<td>OSDUHS</td>
<td>Ontario Student Drug Use and Health Survey</td>
</tr>
<tr>
<td>RMSEA</td>
<td>Root Mean Square Error of Approximation</td>
</tr>
<tr>
<td>SRMR</td>
<td>Standardized Root Mean Square Residual</td>
</tr>
<tr>
<td>WLSMV</td>
<td>Weighted Least Square Mean and Variance Adjusted</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction and Background

1.1 Purpose

This thesis was conducted (and partially funded) under the auspices of the Opening Minds anti-stigma initiative of the Mental Health Commission of Canada as part of its then, 10-year mandate to reduce stigma in Canada. The approach taken by the Opening Minds initiative incorporated a grassroots community development philosophy with clearly defined target groups, in this case adolescents, with contact-based education as the key intervention to reduce stigma (prejudice and discrimination). Contact based education programs share the goal of reducing mental health related stigma by focusing on people with lived experiences of a mental illness telling their personal recovery stories to students who then have an opportunity to ask questions and engage in active discussion. A complete description of this initiative is contained elsewhere (Stuart, et al., 2014a; Stuart, et al., 2014b).

The purpose of the research was to work with a network of community programs that were actively engaged in anti-stigma work targeting youth using contact-based interventions in order to evaluate their effects in reducing stigma, defined as improving stereotypic attitudes and social acceptance toward people with a mental illness. Within this research stereotypes are negative expectations about a person with a mental illness (e.g., a person with a mental illness is unreliable and dangerous). Most public stigma studies assess some form of stereotype awareness, for example, dangerousness, unpredictability, or incompetence (Corrigan, et al., 2001; Martin, Pescosolido, & Tuch, 2000). Social acceptance is typically defined in terms of social distance. Bogardus (1925) originally examined social distance in relation to social acceptance of different racial groups. In 1957 a Canadian research team, Cumming and Cumming (1957), were the first to use social distance to measure change in public tendencies to
discriminate against people with a mental illness. Most studies looking at stigma towards the mentally ill use some variant of Bogardus’s social distance construct to assess social acceptance. It reflects an individual's self-report on relative willingness to participate in different types of relationships of varying degrees of intimacy, with a person who has a stigmatized identity (Corrigan, Edwards, Green, Diwan, & Penn, 2001; Link, Yang, Phelan, & Collins, 2004). The desire for social distance towards people with a mental illness is considered as one of the indicators for social intolerance and discrimination (Dietrich, Beck, Bujantugs, Kenzine, Matschinger, & Angermeyer, 2004). Stereotypic attitudes and social acceptance were the aspects of stigma that were targeted for change by the contact-based anti-stigma programs under study.

1.2 Objectives:

1.2.1 Scale Development

In their review of the literature of measures of mental illness stigma, Link et al. (2004) found research focused on adolescents accounted for only 4% of all research on mental illness stigma (Link, Yang, Phelan, & Collins, 2004). The review was designed to help researchers select measures, but none were specifically developed and psychometrically tested for adolescents. Some measures have been developed for specific adolescent studies (Rahman, Mubbashar, Gater, & Goldberg, 1998; Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004), but often they have poor or untested reliability. Also, although there are some commonly used instruments to measure mental illness stigma for adults, such as the Perceived Devaluation-
Discrimination Scale (Link, Cullen, Frank, & Wozniak, 1987), the Attribution Questionnaire (Corrigan P., et al., 2002; Corrigan, Markowitz, Watson, Rowan, & Kubiak, 2003) and the Opinions about Mental Illness Scale (Cohen & Struening, 1962), there is limited evidence to support reliability and validity in adolescents (Link et al., 2004). Measures for adults have also been adapted for stigma assessment among adolescents and children. The Attribution Questionnaire was revised (Watson, et al., 2004) and is referred to as the Revised Attribution Question. It has been used in adolescent populations (Corrigan, Lurie, Goldman, Slopen, Medasani, & Phelan, 2005; Corrigan P., et al., 2007), but psychometric data for adolescents have been limited to reporting of reliability. Individual survey items, but not the instrument as a whole, have been used to represent a single attribute and have been examined in testing of hypotheses (Corrigan P., et al., 2007; Corrigan, Lurie, Goldman, Slopen, Medasani, & Phelan, 2005). Because we could not identify any appropriate scales from the literature, the first objective of this research was to develop and psychometrically test two scales that could be used to assess stigma reduction in the programs partnering with this evaluation initiative. The first scale was to measure knowledge and stereotypic attributions towards those with a mental illness, and the second was to measure social acceptance of those with a mental illness. These were the two aspects of stigma that the program partners were targeting with their educational approaches.

1.2.2 Outcome Evaluation

Once scales were developed and tested, the second objective was to evaluate the extent to which partner programs were successful in reducing stigma in their respective target populations using variants of the contact-based approach. We used a utilization focused orientation in
conducting this evaluation because it has been developed specifically to assist stakeholders in program decision-making. Utilization-focused evaluation is an approach which tries to maximize the potential for action by identifying the intended uses of an evaluation in the planning stages. It is a decision-making framework for enhancing the utility and actual use of evaluations. It begins with the premise that evaluations should be judged by their utility and actual use, the focus is on intended use by intended users. Therefore, evaluators facilitate the evaluation process and design any evaluation with careful consideration of how everything that will be done, from beginning to end, will affect use. (Patton M. Q., 2008). A utilization-focused evaluation does not require particular evaluation content, model or method, or a particular kind of use. It can include a wide variety of evaluation methods. It is a process for making decisions about issues in collaboration with an identified group of primary users focusing on their intended uses of the evaluation results.

Patton argues that research on evaluation demonstrates that intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings and they are more likely to understand and feel ownership if they've been actively involved (Patton M. Q., 2008). Intended users may be more likely to act upon data when (1) data answer the questions to which they want answers; (2) data are credible to them; (3) data are reported in a concise, understandable manner; (4) data are delivered before decisions need to be made; and (5) stakeholders make time to reflect on findings, implications, and possible actions (Patton M. Q., 2008). There are two essential elements in this approach. First, the primary intended users of the evaluation must be clearly identified and personally engaged at the beginning of the evaluation process to ensure that their primary intended uses can be identified. Second, evaluators must
ensure that these intended uses of the evaluation by the primary intended users guide all other decisions that are made about the evaluation process. To promote a utilization focused evaluation we followed Patton’s (2008) five core objectives: (1) clearly identifying the stakeholders who will use the evaluation findings to make informed decisions about the program; (2) actively including decision-makers in every aspect of the program evaluation in a manner that increases their commitment to utilization of the findings; (3) actively educating all stakeholders in the nuances of effective program evaluation so they are capable of playing an active key role in methodological decisions; (4) presenting genuinely useful findings to decision makers to reinforce future commitment to evaluation; and (5) strategically embedding the standards of utility, feasibility, propriety and accuracy (Yarbrough, Shulha, Hopson, & Caruthers, 2011). We recognized that utilization focused evaluations are influenced by real-world constraints and worked extensively with program partners to ensure that methods were practical and feasible (i.e., conducted within the confines of resources, time, and context), served a useful purpose (produced findings that could be used to improve program delivery), and conducted in an ethical manner with buy-in from all participants. This included a large knowledge exchange component.

1.3 Thesis Organization

This thesis consists of six chapters including this introductory chapter. The second chapter includes a literature review focusing on the conceptualization of mental health stigma, stigma reduction strategies, contact theory, a review of existing measures, and considerations to keep in mind when developing scales and a framework for utilization focused approach. The third chapter is written in manuscript format and is the development and psychometric analysis of two
scales to measure youth attitudes and social acceptance of the mentally ill. The fourth chapter is written in manuscript format as a practice note and describes a utilization focused evaluation of contact-based stigma reduction programs directed toward youth. The fifth chapter is written in manuscript format and is an analysis of a non-randomized comparison group design looking at the effect of three different contact based education interventions. The sixth chapter is a discussion of the findings from this work and a consideration of the next steps to continue this work.
Chapter 2 – Background

This chapter provides a brief background on the nature of stigma, the limited data on what we know about the epidemiology of stigma in youth, and the theory and research behind contact-based stigma reduction programs. Background literature on selected methodological considerations that are not contained in the manuscripts (because of length restrictions) are also provided. The background literature on scale development and program outcomes are incorporated into the manuscripts that are presented in subsequent chapters.

2.1 Mental Illness Stigma

Mental illness stigma occurs when individuals are devalued or treated unfairly by others because of their mental health condition. In Canada and elsewhere, the stigmatization of people with mental illnesses has been recognized as an important public health and human rights problem. The World Health Organization has identified stigma as a "hidden burden" of mental illness and a major public-health challenge (World Health Organization, 2001). International conventions, such as the United Nations Convention on the Rights of Persons with Disabilities, have challenged signatories, of which Canada is one, to take all appropriate measures to eliminate discrimination on the basis of disability. This includes raising awareness and eliminating stereotypes, prejudices and harmful practices (United Nations General Assembly, 2006).

There is no single definition or theoretical perspective on stigma. Present views are rooted in social and psychological traditions, resulting in terms being used in different ways
depending on the perspective. Goffman (1963), a sociologist, first defined stigma as “a dynamic process of devaluation that significantly discredits an individual in the eyes of others.” He viewed stigma as a process based on the social construction of identity. Persons who become associated with a stigmatized condition thus pass from a “normal” to a “discredited” or “discreditable” social status (p. 3). Goffman examined various forms of stigma but recognized that mental illnesses were amongst the most deeply discrediting of all stigmatized conditions and that people with a mental illness were irredeemably marked or tainted.

Psychological theorists have focused on cognitive and attributional processes that lead to the development and maintenance of negative and erroneous stereotypes. Attribution theory describes a relationship beginning with a label, which triggers stereotyped attributions. The attribution may then evoke a negative emotional response (prejudice) and a behavioural expression (discrimination) (Hindshaw & Stier, 2008). For example, when an individual’s illness is attributed to forces within their control, the person is likely to be held responsible for their condition, and subsequently stigmatized. Research has shown that people who hold moral models of mental illness based on attributions -- such as people with a mental illness are to be blamed for their symptoms or that the illness is controllable -- are more likely to respond in an angry and punitive manner and are more likely to be socially intolerant and distancing (Corrigan, Markowitz, Watson, Rowan, & Kubiak, 2003).

Building on these ideas Link and Phelan (2001) provide the most comprehensive conceptualization of stigma. Their conceptualization of stigma, which is commonly used in the literature, involves several interrelated elements that draw attention away from individual characteristics and emphasize social structures. The first element involves the identification and labelling of difference by the social group. This is followed by cultural beliefs that link the label
and the labelled individuals to negative stereotypes. For many persons, the “mentally ill” label is linked to the stereotypes of dangerousness, violence, poor prognosis, disruption of social interaction, and attribution of responsibility (Hayward & Bright, 1997). Next, a division is made between “them” and “us” where “them” are the individuals identified as different and to whom the stereotype is applied. Once categorized in this way, they experience status loss and discrimination. Finally, stigmatization is seen as depending on social structures that provide unequal access to social, economic, and political power. From this perspective, stigmatization is the result of a complex social process involving multiple and mutually reinforcing elements.

A typical psychological model, such as Corrigan and Watson’s (Corrigan & Watson, 2002), puts more stress on the cognitive-emotional-behavioral core features of stigmatization. Link and Phelan’s social structural model is based more on societal aspects such as labeling theory and social power structures. However, in Link and Phelan’s model, stereotypes, separation, and status loss/discrimination parallel Corrigan and Watson’s stereotype, prejudice, and discrimination (Rüsch, Angermeyer, & Corrigan, 2005). Figure 1 shows a side by side comparison of these two models.
In 2008, the United Nations Convention on the Rights of Persons with Disabilities (to which Canada is a signatory) came into force (United Nations General Assembly, 2006). It promotes, protects, and ensures the full and equal enjoyment of human rights and fundamental freedoms of people with disabilities, including people with mental and intellectual disabilities. Signatories agree to adopt appropriate legislative, administrative, and other measures to ensure that the rights of people with disabilities are recognized. The Convention also requires parties to
raise awareness about the rights of people with disabilities, foster respect, and combat stereotypes, prejudices, and harmful practices. The Mental Health Commission of Canada’s *Opening Minds* initiative, through its anti-stigma activities, helps Canada meet our international obligations under the Convention. The Convention provides a legal framework for putting an end to discrimination experienced on a daily basis by people with mental disabilities. Thus, finding strategies for changing stigma has become an important area of research.

### 2.2 Epidemiology of Stigma in Youth

The Mental Health Commission of Canada has identified youth as an important target for anti-stigma activities as they are at high risk of being stigmatized because of a mental health related issue. The onset of many mental illnesses occurs during childhood or adolescence. Fifty percent of people with a mental illness will experience its onset by age 11, and 75% before the age of 25 (Kessler, Berglund, Demler, Jin, Merikangas, & Walters, 2005). Mental illnesses are considered to be the most prevalent source of disability in adolescents. It has been estimated that 14% to 25% of youth have some form of mental illness (Waddell, Offord, Shepherd, Hua, & McEwan, 2002; Waddell, Shepherd, Chen, & Boyle, 2013; Gore, et al., 2011; Health Canada, 2002). The number of Canadian youth living with a mental illness was estimated to be over 1 million in 2011 and it is estimated that, by 2041, there will be almost 1.2 million children and adolescents between the ages of 9 and 19 living with a mental illness in Canada (Smetanin, P; Stiff, D; Briante, C; Adair, C; Ahmad, S; Khan, K;., 2011). There is also some evidence that Canadian youth may experience higher levels of emotional distress than youth in other countries. In a multi-country study conducted by the World Health Organization, Canadian students were among the most likely to report feeling depressed for a week or more, with estimates ranging
from a quarter to over one third, depending on age and gender (World Health Organization, 1996). Thirty-seven percent of high school students in one Southeastern Ontario school district reported multiple symptoms of emotional distress (such as depression or anxiety) and 62% reported multiple stressors from school, work, parents, and friends (Stuart H., 2006). The Centre for Addiction and Mental Health’s Ontario Student Drug Use and Health Survey (OSDUHS) found that about one-fifth (21.9%) of students in grades 7 through 12 reported visiting a professional about a mental health issue at least once in the past year. This estimate represents about 227,500 students in Ontario. This was a significant increase from 2011 (15.1%) and 1999 (12.4%). The same study found that 15.3% of Ontario students rate their mental health as fair or poor. This estimate represents about 157,900 students in Ontario. The percentage of students rating their mental health as fair/poor in 2013 (15.3%) did not significantly differ from 2011 (13.7%). However, the 2013 percentage was significantly higher than that found in 2007 (11.4%), the first year of monitoring (Boak, Hamilton, Adlaf, Beitchman, Wolfe, & Mann, 2014).

Despite the high prevalence of mental disorders among adolescents, they are the least likely age group to seek help (Polanczyk, Salum, Sugaya, & Rohde, 2015). One of the most significant barriers to seeking help and continuing with treatment is the stigma associated with mental illnesses (Angermeyer & Dietrich, 2006; Boyd, Katz, Link, & Phelan, 2010; Corrigan, Druss, & Perlick, 2014; Rüsch, Angermeyer, & Corrigan, 2005). Fear of stigmatization is one of the reasons reported by young people as affecting their willingness to seek help for a mental illness (Corrigan P., 2004; Chandra & Minkovitz, 2007; Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004; Bowers, Manion, Papadopoulos, & Gauvreau, 2013; Gulliver, Griffiths, & Christensen, 2010). Many adolescents worry about others viewing them negatively, and fear
being judged as weak (Chandra & Minkovitz, 2007). In addition, youth who do receive treatment are more likely than older respondents to report being stigmatized. Research examining a national sample of Canadians reported youth (age 12 to 25) who received treatment for a mental health problem in the year prior to the survey were more likely than older respondents to report being stigmatized as a result of a current or past mental or emotional problem (Stuart, et al., 2014a). These results support the need for anti-stigma programs that target youth.

2.3 Stigma Reduction

Although much research has been devoted to the development and origins of stigma and its consequences, less is known about what makes an anti-stigma program successful. Therefore, identification and evaluation of effective methods for intervention are important to the development of national programs to reduce stigma (Stuart, Arboleda-Florez, & Sartorius, 2012).

Work to reduce the stigmatization associated with mental illnesses has been carried out in many ways and the literature currently identifies six main strategies. These are education, protest, contact, legislative reform, advocacy and stigma self-management (Corrigan, et al., 2001; Arboleda-Flórez & Stuart, 2012). Education aims to replace myths and stereotypes about mental illness with correct information (Corrigan P. W., 2004). Protest uses strategies to demonstrate and challenge social injustices (Larson & Corrigan, 2008). Contact reduces stigma using direct interactions with people who have experienced a mental illness. Couture and Penn (2003) describe contact as an attempt to dispel inaccurate and negative beliefs about mental illnesses by placing people in direct personal contact with the stigmatized group. Legislative reform prohibits discrimination, improves protections, and offers accommodations (in areas such as employment, education and housing) for people with a mental illness. Advocacy is used to
ensure people with a mental illness actually realize the rights and freedoms offered by legislation. Stigma self-management programs and supports aim to empower people with a mental illness to overcome their illness identity and find new personal meaning and social roles with the recognition that recovery means more than symptom control and disease management. Although there is no practical way to compare the six strategies for effectiveness, a meta-analysis conducted by Corrigan and colleagues (2012), which included protest or social activism, education of the public, and contact with persons with mental illness, concluded that contact had a better effect on attitudes in adults whereas education had a better effect on attitudes in adolescents. It was suggested that beliefs about mental illnesses might not be as firmly developed in adolescents as they are in adults, making adolescents more likely to be responsive than adults to education. However, the majority of existing evidence supports contact-based education as a promising strategy for improving knowledge, attitudes, and behavioural intent toward mental illnesses (Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004; Sakellari, Leino-Kilpi, & Kalokerinou-Anagnostopoulou, 2011; Stuart H., 2006; Wei, Hayden, Kutcher, Zygmunt, & McGrath, 2013; Kolodziej & Johnson, 1996; Pinfold, Stuart, Thornicroft, & Arboleda-Flórez, 2005).

2.4 Social Contact Theory

The effect of contact on attitudes towards stigmatized groups was formally conceptualized as the contact hypothesis by Gordon Allport in his 1954 book *The Nature of Prejudice* (Allport, 1954). Prejudice, the emotional/attitudinal component of stigma, is seen as a direct result of generalizations and oversimplifications made about a group of people based on incomplete or mistaken information. Although Allport’s target was to improve intergroup
relations between different races, his theory has since been extended to other target groups including the mentally ill.

Allport specified four conditions for optimal intergroup contact to counter negative stereotypes and reduce prejudice. The groups should have equal status within the contact situation, they should work on common goals, they should cooperate with each other, and there should be support from authorities (institutional support) (Allport, 1954). Pettigrew added a fifth condition, the contact should carry a potential for the formation of friendships with members of the other group (Pettigrew, 1998; Pettigrew & Tropp, 2006). The contact hypothesis has traditionally held that Allport’s optimal conditions are essential. However, Pettigrew and Tropp (2006) conducted a meta-analysis and found that although studies designed to meet Allport’s optimal conditions showed stronger contact–prejudice effects, the presence of all four conditions was not essential for intergroup contact to achieve positive outcomes.

The original contact hypothesis says nothing about the processes by which contact changes attitudes and behavior. It predicts only when contact will lead to positive change, not how and why the change occurs. Several mechanisms have been proposed to explain how contact can reduce prejudice. Pettigrew (1998) attempts to address the question of process (the how and why) through four interrelated processes that operate through contact: learning about the out-group, changing behavior; generating affective ties; and in-group reappraisal. Contact works through cognitive, behavioural and affective means. The cognitive could include such things as learning about the out-group or reappraising how one thinks about one’s own in-group. The behavioural would be changing one’s behaviour to open oneself to potential positive contact experiences. The affective or emotional would be generating affective ties and friendships and
reducing negative emotions (anxiety / threat) (Pettigrew & Tropp, 2008; Tausch & Hewstone, 2010).

In this study, contact could be direct (in person), by video (indirect), or by other vicarious means such as theatre or books, all of which centered on a person or character with personal experience of a mental illness.

Based on the foregoing literature, it is important to recognize that most major mental disorders have their onset in adolescence, making this an important time for early identification and treatment. However, stigma and stigma avoidance can undermine our ability to intervene early and minimize psychosocial and occupational impairments. Contact-based education may be a means to reduce mental health related stigma and may make it easier for students to ask for help.

The next section provides a narrative review of the methodological literature that is background to this thesis not covered in detail in any of the manuscripts. It includes a discussion of instrument development and testing, and more detailed analysis of the program evaluation approach used.

2.5 The nature of the interventions under study

To more thoroughly understand the programs, a separate qualitative study was carried out to build a program logic model to indicate how the various parts of contact-based programming interact to produce an effect (figure 2). The research team worked with 18 contact-based
educational programs implemented across Canada that targeted high school students, all formally affiliated with the *Opening Minds* anti-stigma initiative.

Research related to critical ingredients and program fidelity testing was reviewed and revealed three major steps for developing a program model (Holter, Mowbray, Bellamy, MacFarlane, & Dukarski, 2004; McGrew & Bond, 1995; Mowbray, Holter, Teague, & Bybee, 2003; Solomon, 2004). The first step involved collecting program information using various methods, including: (1) a review of previous research and published descriptions of existing programs that have some evidence of positive outcomes; (2) gathering stakeholder opinions by interviewing people who have implemented or received the program, either individually or in focus groups; (3) site visits and in-person observations of a particular model program, and (4) getting consensus from expert panels. The second step involved constructing a list of critical ingredients based on the program information collected in the first step and then grouping the ingredients into domains for conceptual clarity. Finally, the third step involved validating the critical domains, usually by having experts or other stakeholders rate the importance of each domain. A more detailed account of the method used can be found in Chen et al (2016).
The inputs refer to the resources invested in a program that are necessary to form a supporting infrastructure for program activities. These include having secure funding, a strong team, good relationships with schools, good speaker preparation and coordination of activities. Processes are the activities of the program in relation to the intervention delivered. These include things such as who delivers it, what they say and do, how information is provided, whether materials are also provided, where the intervention is provided (e.g. a class-room setting, assembly, or summit), and when. Finally, the outcomes define the expected program effects. In this logic model, the outcomes are based on a series of levels, ranging from the simplest (student reaction to the program) to the most difficult to achieve (social changes). This logic model was developed to help guide program activities along some logical course in order to maximize their chances of success.

The quantitative data used for this thesis were collected from students in grades 8 through 12 between March 2010 and December 2012. Twenty programs delivering 29 interventions across seven provinces and territories (Alberta, Manitoba, Saskatchewan, Ontario, Quebec, Nova Scotia and Northwest
Territories) collected 11983 baseline surveys. Table 1 gives the sample size for each analysis. The test-retest sample and the samples used in manuscript three were not unique in that they were also included in the sample used in manuscript two.

**Table 1 Sample Size**

<table>
<thead>
<tr>
<th>Manuscript 1</th>
<th>Number of Interventions</th>
<th>Baseline N</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploratory factor analysis</td>
<td>6</td>
<td>1352</td>
<td>2010</td>
</tr>
<tr>
<td>Test - retest</td>
<td>1</td>
<td>233*</td>
<td>2011-2012</td>
</tr>
<tr>
<td>Confirmatory factor analysis</td>
<td>4</td>
<td>576</td>
<td>2011-2012</td>
</tr>
<tr>
<td>Manuscript 2</td>
<td>Program evaluation</td>
<td>29</td>
<td>11983</td>
</tr>
<tr>
<td>Manuscript 3</td>
<td>Summit</td>
<td>1</td>
<td>106*</td>
</tr>
<tr>
<td>In-class</td>
<td>1</td>
<td>1212*</td>
<td>2011-2012</td>
</tr>
<tr>
<td>Assembly</td>
<td>1</td>
<td>467*</td>
<td>2011-2012</td>
</tr>
<tr>
<td>Comparison group</td>
<td>1</td>
<td>361*</td>
<td>2010-2013</td>
</tr>
<tr>
<td>Unique sample</td>
<td>29</td>
<td>13983</td>
<td>2010-2012</td>
</tr>
</tbody>
</table>

*Samples were not unique in that they were also used in the program evaluations*
2.6  Review of Instrument Development

Measurement scales are “collections of items combined into a composite score, and are intended to reveal levels of theoretical variables not readily observable by direct means” (Devellis, 2003, p. 8). There are various strategies used in scale construction. Basic principles of scale development have been described by a number of authors (DeVellis, 2003; Streiner & Norman, 2008; Clark & Watson, 1995) . These authors suggest that the process of scale development should include the following steps: 1) conceptualization and development of an initial item pool; 2) item selection and psychometric evaluation, and 3) the ongoing process of validation. However, even before this you have a well-developed model or theory about the construct to be measured such as the one described by Link and Phelan (Link & Phelan, 2001) previously described.

2.6.1  Item Generation and Scale Construction

Streiner and Norman (Streiner & Norman, 2008) suggest five sources for new items. They can come from theory, the subjects themselves, observation, research, and expert opinion. They recommend that the first step in devising items is to review the literature to see what others have done. This can be followed by focus groups and expert opinions to help in the selection and refinement of items, and to generate items to fill in any gaps.

Once items have been developed, a questionnaire needs to be constructed. There are a number of general principles to consider to ensure each question is clear and understood, the response options are appropriate, and that the layout and organization of the questionnaire enables easy completion. These are addressed where they apply in the manuscript in chapter 3.
2.6.2 Validity and Reliability

2.6.2.1 Validity

One of the overall goals when constructing scales is that they display good psychometric properties (reliability and validity) and can be used to make inferences in a wide range of samples. Within this context, validation refers to "the inferences that can be made about the attributes of people who have produced those test scores" (Landy, 1986, p. 1186). That is, according to Streiner and Norman, validity is not an attribute of the test, but "is really a process whereby we determine the degree of confidence we can place on inferences we make about people based on their scores from that scale" (2015, p. 230). Criterion, divergent, and discriminant validation are all different aspects of validity. As stated in the Standards for Educational and Psychological Testing: "The sources of evidence may illuminate different aspects of validity, but they do not represent distinct types of validity. Validity is a unitary concept. It is the degree to which all of the accumulated evidence supports the intended interpretation of the test scores for the intended purposes." (1999, p. 11).

Content validation includes strategies that focus on the content of the scale. One aspect of content validation is assessing whether the instrument appears to measure what is intended to measure. Content validation can be addressed with the use of experts and theory (Nunnally & Bernstein, 1994).

Construct validation traditionally has been defined as the experimental demonstration that a test is measuring the construct it claims to be measuring. Exploratory factor analysis can be informative in this respect, but confirmatory factor analysis is required to fully understand whether the scale is measuring the underlying constructs it is hypothesized to be measuring.

Criterion validation examines the extent to which results from the questionnaire are consistent with a previously developed test that is considered the gold standard. However, in
many cases, there is no gold standard available, or it is too costly or distant in time to assess (Streiner D., 2003; Streiner & Norman, 2008). DeVellis (2003) suggests known group validation can help support criterion validation. Known group validation is sometimes referred to as discriminative validation (Cappelleri, Zou, Bushmakin, Alvir, Alemayehu, & Symonds, 2013; Streiner & Norman, 2015). It involves demonstrating that a scale can differentiate one group from another based on their scale scores. If the measure is able to discriminate between the groups through statistically significant findings, this provides evidence for the validity of the measure. All of these approaches were used in this research to assess the validity of the newly developed scales.

2.6.2.2 **Reliability – Internal Consistency**

Reliability refers to the repeatability, stability or internal consistency of a scale. One of the most common ways to demonstrate this uses the Cronbach’s alpha statistic. This statistic uses inter-item correlations to determine whether items are measuring the same domain. Cronbach’s alpha reliability coefficient normally ranges between 0 and 1. The closer Cronbach’s alpha coefficient is to 1.0 the greater the internal consistency of the items in the scale. The size of alpha is determined by both the number of items in the scale and the mean inter-item correlations. Cronbach’s alpha of 0.70 has been suggested as is an acceptable (minimum) level of reliability for exploratory studies. It also has a suggested upper limit of about 0.90, above that, there is too much redundancy in the scale (DeVellis, 2003; Nunnally & Bernstein, 1994). Item-partial total correlation (the correlation between each item and the remaining scale score) can be used to assess internal consistency. If the items are measuring the same underlying concept, then each item should correlate with the total score from the questionnaire or domain. Kline (1993) and
Nunnally and Bernstein (1994) recommend deleting any questionnaire item with a corrected item-total correlation of <0.3. Item analysis using inter-item correlations will also identify those items that are too similar. High inter-item correlations (>0.8) suggest that these are repetitions of each other, and they should also be removed. The alpha-if-item-deleted is used to determine the scale’s alpha if the item were to be deleted from the scale. It is important for each item to contribute to the scale’s alpha. If an item does not contribute to the scale’s alpha then it should be considered for deletion (Kline, 1993).

2.6.2.3 Reliability – Test-Retest

Another important aspect of reliability is the stability of the scales over time. Theoretically, if a measure truly reflects a construct, assessment of that construct should be comparable on different occasions. Better reproducibility suggests better precision of measurements, which is a requirement for better tracking of changes. Various statistical methods have been used to test for agreement. Which method is the best is still open to debate and almost all methods have been criticized. One approach for measuring agreement is the correlation coefficient. However, this only measures the strength of linear association between variables. Comparing means has also been suggested for assessing agreement. Bland and Altman plots are another method that can be used. In a Bland and Altman plot, the difference between the two measurements per subject is plotted against the mean of the two measurements. The limits of agreement are then defined as −1.96 s and +1.96 s, with s being the observed standard deviation of the difference between the two measurements per subject (Bland & Altman, 1986).

However, the favorite method is the intraclass correlation. The ICC also eliminates the problem of measuring “only” linear association, because there’s a penalty if there is any systematic bias.
2.6.2.4 Validity - Confirmatory Factor Analysis

If the above steps are all carefully followed, it is highly likely that the new scale will be internally consistent and possess content validity. One of the weaknesses of typical factor analytical techniques is their inability to quantify the goodness of fit of the resulting factor structure. At this stage it is recommended that confirmatory factor analysis be conducted on data collected from an independent sample (Worthington & Whittaker, 2006). The major advantage of CFA is that a structure is specified a priori; hence, it is testing a hypothesis, which EFA does not do. Confirmatory factor analysis also allows for the quantitative assessment of the quality of the factor structure providing further evidence of the construct validity of the new measure. In this thesis, both exploratory and confirmatory factor analysis (using a separate sample as recommended) are conducted.

2.6.3 Sample Size Considerations

2.6.3.1 Exploratory Factor Analysis

Although sample size is an important consideration in factor analysis, there are varying opinions about how to determine sample size with several guiding rules of thumb cited in the literature. For example, Kline’s absolute minimum of 100 (1993) and Comrey and Lee’s (1992) rule of 500 (following the guide 50 as very poor; 100 as poor, 200 as fair, 300 as good, 500 as very good and 1000 as excellent). Another set of recommendations has been made in terms of the ratio of sample size to the number of variables. Cattell (1978) recommended a ratio ranging from 3:1 to 6:1; Everitt (1975) proposed a ratio of at least 10:1; Hair et al. (1995) suggested a
ratio of 20:1; whereas Nunnally’s suggested a ratio of at least 10 participants to each variable (Nunnally, 1978). If there are missing values, sample size must be considered as well as if the missing values occur in a nonrandom pattern. It has been recommended that cases with missing values be deleted to prevent overestimation (Tabachnick & Fidell, 2007).

### 2.6.3.2 Confirmatory Factor Analysis

One concern when conducting CFA is the effect of the sample size on the goodness of fit indices. According to Bryant and Yarnold (Bryant & Yarnold, 1995) the chi-square statistic is sensitive to sample size and thus small samples may provide less power to detect a model’s true lack of fit. However, there are no absolute recommendations when it comes to the adequate sample size to obtain reliable results (Bryant & Yarnold, 1995; Tabachnick & Fidell, 2007). A common rule of thumb for determining sample size is having a minimum of 200 observations (Shar & Goldstein, 2006). Recommendations for absolute sample sizes vary from a minimum of 50 participants to 300 or more, while other offer recommendations in terms of ratios such as a five-to-one or a 20-to-one ratio of participants-to-variables (Furr & Bacharach, 2008).

### 2.7 Program Evaluation

#### 2.7.1 Research and Program Evaluation

The terms research and evaluation are sometimes referred to synonymously, but are different in their purpose. “Even though evaluation and research provide information about questions of interest, they are typically undertaken for different reasons” (Russ-Eft & Eft, 2009, p. 9). Research is a formal and systematic application of the scientific method to the study of problems
(Gay, Mills, & Airasian, 2008). Research is seen as more interested in producing generalizable knowledge, more theoretical, more controlled by the researchers, while evaluation is seen as more interested in specific, applied knowledge, and more controlled by those funding or commissioning the evaluation. Spaulding (Spaulding, 2008) argued that although a research study could examine a particular program, most research tends to be interested in either generalizing findings back to wider audiences or discussing how the study’s findings relate back to the literature. Evaluation, on the other hand, is a systematic collection of information about activities, characteristics, and or outcomes of programs in order to make judgments about the program, to improve its effectiveness, and/or to inform decisions about future programming (Patton M. Q., 2008). It is a form of inquiry that seeks to address questions concerning how well a program, process, product, system, or organization is working. It is typically undertaken for decision-making purposes and should lead to a use of findings by a variety of stakeholders (Russ-Eft & Eft, 2009, p. 6). Evaluation is the process of determining the merit, worth, and value of things, and evaluations are the products of that process (Scriven, 1991, p. 139). Because it is program specific, it may not provide generalizable knowledge, as would be the goal of research.

2.7.2 Program Evaluation

The literature on evaluation practice, evaluation use, and the factors that promote use is immense. The term evaluation means different things to different people (King & Stevahn, 2013). Program evaluators come from a wide range of disciplines and professions with different orientations, methods and motivations, and this has contributed significantly to the many perspectives. Many in the field have pointed out that there really is no single definition that adequately captures all of this diversity (Kahan, 2008). A widely referred to definition is that of Patton (2008): “Evaluation is the systematic collection of information about the activities,
characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding” (p. 39).

Evaluation has been classified into five types by intended use: formative, process, summative, outcome, and impact. Formative evaluation provides information to guide program improvement, whereas process evaluation determines whether a program is delivered as intended to the targeted recipients (Rossi, Lipsey, & Freeman, 2004). Formative and process evaluations are appropriate during the implementation of a program. Summative evaluation informs judgments about whether the program worked (i.e., whether the goals and objectives were met) and requires making explicit the criteria and evidence being used to make “summary” judgments. Outcome evaluation focuses on the observable conditions of a specific population, organizational attribute, or social condition that a program is expected to have changed. Whereas outcome evaluation tends to focus on conditions or behaviors that the program was expected to affect most directly and immediately (i.e., “proximal” outcomes), impact evaluation examines the program’s long-term goals. This thesis undertook outcome evaluations of contact-based educational programs offered in middle and high schools in various locations across Canada. However, previously, we had conducted a process evaluation to understand the logic underlying the program and present the results of the logic model to help situate the programs under investigation. (Chen, Koller, Krupa, & Stuart, 2016).

Many evaluation approaches and models exist and the literature suggests no one approach is best for all situations, as each approach has its own set of strengths and weaknesses. For example, Kahan (2008) separates the major evaluation approaches into results focused (goal...
based, goal free, theory based), utilization focused, collaborative, balanced scorecard and appreciative inquiry. The best approach varies depending on factors such as fit with basic values, the intent of the evaluation, the nature of key stakeholders, and the available resources. In addition, it is not necessary to stick strictly to one approach; evaluations might combine elements of different approaches (Rogers & Fraser, 2003).

2.7.3 Evaluation Standards

Because research and evaluation have different objectives, they should be evaluated by different standards (Patton M. Q., 2008). A set of evaluation standards were established in 1981 by the Joint Committee on Standards for Educational Evaluation and were most recently revised in 2011. The standards are used to address the quality of evaluation efforts (Yarbrough, Shulha, Hopson & Caruthers, 2011). These standards are often cited and have been approved by the Canadian Evaluation Society, the Canadian Society for the Study of Education, the American National Standards Institute, and they have also been endorsed by many other professional organizations. These standards do not define a way to do evaluation; rather, they serve to guide choice from among the many options available at each step in the Framework. The 30 standards are grouped into five categories (utility, feasibility, propriety, accuracy and accountability) which provides direction for many of the decisions made in the progression of an evaluation (See Appendix).

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs. Who will use the results? Will the evaluation provide relevant information in a timely manner for them? Stufflebeam
(2001), carried out an extensive review of 22 evaluation approaches. He identified nine as the strongest and most promising for continued use and development. Utilization-focused evaluation was among the nine, with the highest rating for adherence to the utility standards. The feasibility standards are intended to increase evaluation effectiveness and efficiency. How much time and effort can be devoted to stakeholder engagement? Are the planned evaluation activities realistic given the time, resources, and expertise at hand? The propriety standards support what is proper, fair, legal, acceptable and ethical in an evaluation. Does the evaluation protect the rights of individuals and protect the welfare of those involved? The accuracy standards are intended to assure truthfulness of evaluation findings. Will the evaluation produce findings that are valid and reliable, given the needs of those who will use the results? The accountability, in the context of program evaluation, refers to the responsible use of resources to produce value as a result of the evaluation (Yarbrough, Shulha, Hopson, & Caruthers, 2011).

2.8 Ethical Considerations

This study was approved by Queen’s University Health Sciences and Affiliated Hospitals Research Ethics and where required, by appropriate School Board ethics/research committees. Confidentiality of the participants was assured. Each student was assigned a unique identifier so that pretests and posttest could be matched anonymously. In addition, programs had the option to remain confidential (as in the case of negative findings) and not have their final reports presented on the Mental Health Commission of Canada’s web site. All programs remained confidential in the context of the data analysis for this thesis. Approval letters can be found in Appendix B.
Chapter 3

Manuscript - Youth Attitudes Towards and Social Acceptance of the Mentally Ill: Development and Psychometric Analysis of Two Scales

3.1 Abstract

**Background:** Adolescents have become an important target for contact based anti-stigma programs. Research on mental health stigma reduction in adolescents has been hindered by lack of valid measures to evaluate program interventions. If we are to systematically reduce stigma and improve mental health, we must have the capacity to observe and measure stigma and to evaluate whether or not program interventions are producing the desired effect. The many scales used to measure attitudes towards people with mental illness that exist today are not adequate as they have been developed for the adult population outside of the intervention setting.

**Methods:** Survey data were collected on three samples of adolescence students aged between 12 and 18 years old (total $N=2118$). Qualitative methods, including expert opinion and focus group discussion were used for item generation, item review and refinement, and in general to establish content validity. Quantitative methods, such as exploratory and confirmatory factor analysis, known group differences, Cronbach’s alpha and test-retest reliability were used to support validity and reliability of the scales.

**Results:** Results showed that a one-factor solution was the best fit for the data for each scale. Following exploratory factor analysis, a confirmatory factor analysis was conducted on a different sample. Results highlighted some items with low loadings. Allowing for several
correlated variances resulted in a good fitting model. These results, in combination with the known group differences and the test – retest reliability, support preliminary reliability and validity of the two scales.

**Conclusions:** This study has established initial reliability and validity of a two scales to evaluate youth attitudes towards and social acceptance of those with a mental illness, with the potential for use in intervention development and evaluation. The Stereotypic Attribution Scale contained 11 items measuring stereotypic attitudes towards people with a mental illness. The Social Acceptance Scale contained 11 items addressing social distance and social responsibility.

**Keywords:** Social Stigma; Mental Illness; Social Distance; Stereotyping; Prejudice; Adolescents; Validation; Scale
3.2 Introduction and Background

This work was undertaken as part of the activities of the Mental Health Commission of Canada’s Opening Minds anti-stigma initiative. The overall approach taken by Opening Minds has been described elsewhere (Stuart, et al., 2014a; Stuart, et al., 2014b). Briefly, the approach incorporated a grassroots community development philosophy with contact-based education as the key intervention used by different community programs to reduce stigma (prejudice and discrimination). Stigma is a multi-faceted construct and can be defined as “a social process, characterized by exclusion, rejection, blame or devaluation that results from experience or reasonable anticipation of an adverse social judgement about a person or group” (Martin & Johnston, 2007, p. 8). Contact based education programs share the goal of reducing mental health related stigma by centering on people with lived experiences of a mental illness telling their personal recovery stories to students who then have an opportunity to ask questions and engage in active discussion. They are considered to be one of the most promising practices for improving prejudicial attitudes and discriminatory behaviours toward people with a mental illness (Couture & Penn, 2003; Pinfold, Toulmin, Thornicrof, Huxley, Farmer, & Graham, 2003) and have been successfully used to reduce stigma in small, selected samples of high school students (Stuart H., 2006; Pinfold, Stuart, Thornicroft, & Arboleda-Flórez, 2005).

The Mental Health Commission of Canada has identified youth as an important target for anti-stigma activities as they are at high risk of being stigmatized because of a mental health related issue. Mental illnesses are considered to be the most prevalent source of disability in adolescents. It has been estimated that 14% to 25% of youth have some form of mental illness (Health Canada, 2002; Waddell, Offord, Shepherd, Hua, & McEwan, 2002; Waddell, Shepherd, Chen, & Boyle, 2013; Gore, et al., 2011). There is also some evidence that Canadian youth may
experience higher levels of emotional distress than youth in other countries. In a multi-country study conducted by the World Health Organization, Canadian students were among the most likely to report feeling depressed for a week or more, with estimates ranging from a quarter to over one third, depending on age and gender (World Health Organization, 1996). Thirty-seven percent of high school students in one Southeastern Ontario school district reported multiple symptoms of emotional distress (such as depression or anxiety) and 62% reported multiple stressors from school, work, parents, and friends (Stuart H., 2006). The Centre for Addiction and Mental Health’s Ontario Student Drug Use and Health Survey (OSDUHS) found that about one-fifth (21.9%) of students in grades 7 through 12 reported visiting a professional about a mental health issue at least once in the past year. This estimate represents about 227,500 students in Ontario. This was a significant increase from 2011 (15.1%) and 1999 (12.4%). The same study found that 15.3% of Ontario students rate their mental health as fair or poor. This estimate represents about 157,900 students in Ontario. The percentage of students rating their mental health as fair/poor in 2013 (15.3%) does not significantly differ from 2011 (13.7%). However, the 2013 percentage is significantly higher than that found in 2007 (11.4%), the first year of monitoring. (Boak, Hamilton, Adlaf, Beitchman, Wolfe, & Mann, 2014).

Despite the high prevalence of mental illness among adolescents, they are the least likely age group to seek help (Polanczyk, Salum, Sugaya, & Rohde, 2015). Fear of stigmatization is one of the reasons reported by young people as affecting their willingness to seek help for a mental illness (Corrigan P., 2004; Chandra & Minkovitz, 2007; Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004; Gulliver, Griffiths, & Christensen, 2010). In addition, youth who do receive treatment are more likely than older respondents to report being stigmatized. Research examining a national sample of Canadians reported youth (age 12 to 25) who received treatment
for a mental health problem in the year prior to the survey were more likely than older respondents to report being stigmatized as a result of a current or past mental or emotional problem. These results support the need for anti-stigma programs that target youth (Stuart, et al., 2014a).

Despite the importance of adolescents as a target for stigma reduction programs, validated measures for this age group are few. In their review of the literature of measures of Mental Illness Stigma Link et al. (2004) found research focused on adolescents accounts for only 4% of all research on mental illness stigma (Link, Yang, Phelan, & Collins, 2004). The review was intended to help researchers select stigma measures, but none were reported to be designed and psychometrically tested for use with adolescents. Some measures have been developed for specific adolescent studies (Rahman, Mubbashar, Gater, & Goldberg, 1998; Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004), but they often have poor or untested reliability. There are some commonly used instruments to measure mental illness stigma for adults, such as the Perceived Devaluation-Discrimination Scale (Link, Cullen, Frank, & Wozniak, 1987), the Attribution Questionnaire (Corrigan P., et al., 2002; Corrigan, Markowitz, Watson, Rowan, & Kubiak, 2003) and the Opinions about Mental Illness Scale (Cohen & Struening, 1962) but there is limited evidence to support reliability and validity in adolescents (Link, Yang, Phelan, & Collins, 2004). Measures for adults have also been adapted for stigma assessment among adolescents and children. The Attribution Questionnaire was revised (Watson, et al., 2004) and is referred to as the Revised Attribution Question. It has been used in adolescent populations (Corrigan, Lurie, Goldman, Slopen, Medasani, & Phelan, 2005; Corrigan P., et al., 2007), but psychometric data for adolescents have been limited to reporting of reliability. Individual survey items, but not the instrument as a whole, have been used to represent a single attribute and have
been examined in prediction of hypotheses (Corrigan, Lurie, Goldman, Slopen, Medasani, & Phelan, 2005; Corrigan P., et al., 2007).

For the purposes of this research the focus was on specific aspects of stigmatization that have been the center of much stigma related research; attitudes or stereotypes and social distance as these are the elements most often targeted for change by the contact-based anti-stigma programs in our network. Within the context of this research stereotypes are defined as negative expectations about a person with a mental illness, for example, dangerousness, unpredictability, and incompetence (Corrigan, et al., 2001; Martin, Pescosolido, & Tuch, 2000). Direct observation and measurement of stigmatizing behaviours is complex and is beyond the scope of this study. Social distance is a measure of behavioural intent and considered to be a fair (but not perfect) proxy for discrimination (Dietrich, Beck, Bujantugs, Kenzine, Matschinger, & Angermeyer, 2004). Most studies looking at stigma towards the mentally ill use some form of social distance scale to assess behavioral intent. Social distance reflects an individual's self-report of their willingness to participate in different types of relationships of varying degrees of intimacy, with a person who has a stigmatized identity (Corrigan, Edwards, Green, Diwan, & Penn, 2001; Link, Yang, Phelan, & Collins, 2004). The desire for social distance towards people with mental illness is considered as one of the indicators for individual discrimination (Dietrich, Beck, Bujantugs, Kenzine, Matschinger, & Angermeyer, 2004).

The goals of the *Opening Minds* Youth initiative were to identify existing anti-stigma interventions that used contact-based education, evaluate their results, and then replicate the most promising approaches across the country. The Network included more than 25 interventions from over 20 programs. Programs signed formal partnership agreements with *Opening Minds* to participate in a third-party evaluation of their activities. Most of the contact-based education
programs that were working with the Mental Health Commission’s *Opening Minds* initiative regularly collected pre and post test data, but with different instruments as there was no standard, psychometrically tested instrument to draw on from the literature. Thus, the first step was to develop and field test a standardized measurement instrument that captured the important dimensions of stigma addressed by program partners. The instrument had to be short, feasible to use in school settings, and contain age-appropriate language (e.g.: for a Grade 5 or 6 reading level). The aim of this research was to develop and psychometrically test two scales that could be self-administered by students and used by programs to assess the effectiveness of stigma reduction programs directed towards youth across different settings. More specifically, the objectives were to establish face and content validity, reduce the number of items and assess preliminary construct validity of each scale using exploratory factor analysis, assess reliability of the scales using Cronbach’s alpha, support construct validity using known group differences, assess the reliability of the scales over time with a test-retest analysis, assess convergent and discriminant validity using exploratory factor analysis, and further validate the two scales using confirmatory factor analysis.

3.3 Methods

3.3.1 Study Setting

In order to promote engagement and the uptake of the evaluation measures by the programs, a utilization focused approach was used (Patton M. Q., 2008). Program stakeholders were involved right from the beginning and played a key role in the development of the scales. They were not going to invest a great deal of time and effort into having their programs
evaluated without the expectation that the outcomes would benefit their program, hence the scales needed to reflect program priorities. Though most programs aimed at reducing stereotypic attributions and increasing social acceptance, the extent to which they addressed each differed. Given this, it was agreed that each domain should have its own scale. Also, they are conceptually distinct. Thus, the first scale was intended to measure attitudes and focused on stereotypic attributions pertaining to perceived controllability of the illness, potential for recovery, and potential for violence and unpredictability. The second scale was to measure behavioural intentions related to social acceptance of people with a mental illness, such as desire for social distance and feelings of social responsibility related to mental health issues.

There was no freedom for researchers to modify the interventions the programs delivered or their method of data collection. Programs were using a quasi-experimental one-group pretest posttest design (Barnighausen, et al., 2017). Baseline data from the first sample were used for the exploratory factor analysis, baseline and follow-up data from the second sample were used in the test-retest analysis and baseline data from the third sample were used for the confirmatory factor analysis.

3.3.2 Item Development

First, a pool of items was developed from existing measures related to mental health stigma published in the scientific literature as well as the unpublished measures used by programs in the field (see Appendix). Items were categorized according to whether they measured stereotypic attitudes or feelings of social distance. Related concepts were mapped onto a chart and reviewed by program staff, an advisory committee composed of both those with lived experience of a mental illness and family members, and researchers. Items were chosen to
represent underlying elements of the constructs of interest. As suggested by DeVellis (DeVellis, 2003), items were reviewed for reading level and clarity. The Flesch-Kincaid grade level for each item was calculated and items were reworded to be accessible to a grade 6 reading level. Double barreled items, items that conveyed two or more ideas, were removed. Because respondents may have a tendency to respond with the same answer on every question, some items were reverse-phrased in an effort to avoid response bias.

Our next task was to choose a scoring method that would generate sufficient variance among respondents for subsequent statistical analyses. Although researchers have used 7-point and 9-point scales, Lissitz and Green (Lissitz & Green, 1975) have shown that, with Likert scales, the reliability coefficient alpha increases up to the use of five points, but then it levels off so it is suggested that the new items be scaled using 5-point Likert-type scales. Therefore a 5-point agreement scale with response options ranging from strongly disagree; disagree; unsure; agree; and strongly agree was used. Higher values indicated higher stigma.

To examine content, readability, and interpretation of items, a draft survey was circulated to all of the programs and the advisory committee members for review and comment. In addition, an informal focus group using a convenience sample of youth, was held. Information gathered was used to investigate whether or not items were understood and were of appropriate length and to decide on the best format for the survey. Questions were refined for content and flow. The first iteration of the instrument contained 15 items addressing stereotyped attributions and 17 items assessing social acceptance.

To examine known groups differences when assessing validity, we also included questions on gender, age, grade, and prior contact with someone with a mental illness. Research suggests females are more socially accepting of people who have a mental illness and have less
prejudiced attitudes toward mental health problems then are males (Corrigan & Watson, 2007; Fox, Buchanan-Barrow, & Barrett, 2008). It has also been shown that those who have had prior contact with someone with a mental illness also show more social acceptance (Corrigan, Edwards, Green, Diwan, & Penn, 2001; Corrigan & Watson, 2002; Pinfold, Toulmin, Thornicrof, Huxley, Farmer, & Graham, 2003; Schulze, Richter-Werling, Matschinger, & Angermeyer, 2003). Research also suggests that those who identify people with a mental illness as dangerous, and those who indicate you cannot rely on someone with a mental illness tend to show less social acceptance toward people with a mental illness (Corrigan W., et al., 2002).

The question for prior contact was: “Do you, or does someone you know, have a mental illness? Response options were: No; Uncertain; Yes, a close friend; Yes, a close family member; Yes, someone other than a close friend or family member; and Yes, I do. Upon completion of the development and refinement of the items, we undertook face and content validation.

3.3.3 Participants

Three separate samples were used in the analysis. For the exploratory factor analysis and the internal consistency analysis six contact-based stigma reduction programs from four provinces (Alberta, Saskatchewan, Ontario and Nova Scotia) agreed to participate. The preliminary instrument was administered by program partners to a sample of 1352 high school youth students in grades 8 through 12 during regular program delivery. One program (from Ontario) agreed to participate in the test–retest study. Data were collected from 190 students in grades 10 and 11. The re-test was provided to students up to two weeks after the initial administration. The confirmatory factor analysis was based on different sample of data collected by four programs on 576 students from Ontario.
3.3.4 Ethics

For all studies a passive consent procedure was used where a letter of explanation was sent home to the parents. The parent needed to sign and return a form if they did not want their child to be included in the evaluation. The evaluation design was approved by Queen’s University Health Sciences and Affiliated Hospitals Research Ethics Board and where required, by appropriate School Board ethics/research committees.

3.3.5 Data Management and Analysis

Completed surveys were returned via mail or by courier to Queen’s University, Canada, where they were double entered and analyzed.

Item level distributions were examined, and descriptive statistics were used to look at the distribution of scores, particularly the range, skewness, and standard deviation, and quantify missing values. Before the extraction of the factors the data were screened to assess the suitability for factor analysis. These tests included Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy, (Kaiser, 1974) and Bartlett's Test of Sphericity (Bartlett, 1950). The KMO index, in particular, is recommended when the case to variable ratio is less than 1:5. The KMO index ranges from 0 to 1, with 0.50 considered suitable for factor analysis (Hair, Anderson, Tatham, & Black, 1995; Tabachnick & Fidell, 2007). The Bartlett's Test of Sphericity should be significant (p<.05) for factor analysis to be suitable (Hair, Anderson, Tatham, & Black, 1995; Tabachnick & Fidell, 2007).

Exploratory factor analysis of the 15 stereotypic attributions and the 17 social acceptance items were carried out separately. Data were analyzed using a Maximum Likelihood extraction
with promax rotation. An orthogonal rotation was selected because it was assumed that if more than one factor was identified, the factors would be correlated. As items were ordered, categorical variables, a polychoric correlation matrix was used, instead of the standard Pearson's correlation matrix. Dimensionality was assessed using Kaiser’s (Kaiser, 1960) eigenvalues greater than one rule, scree plots and Horn’s parallel analysis (Horn, 1965; Thompson & Daniel, 1996). The magnitudes of factor loadings were examined, items with loadings less than .40 were removed and the analysis was repeated.

We used Cronbach’s alpha to assess internal reliability of the resulting scales. Cronbach’s alpha of 0.70 has been suggested as is an acceptable (minimum) level of reliability for exploratory studies (Streiner D., 2003).

As previously indicated, one program agreed to participate in the test-retest. The re-test was provided to students up to two weeks after the initial administration. Surveys were completed anonymously, with test and retest surveys being matched using unique identifiers. Test retest reliability was assessed using Bland-Altman plots. Plots were created for each scale with 95% limits of agreement. First one sample t-tests were run to test whether the mean change between scales scores was significant. If not significant, difference scores were plotted against average scale scores and 95% confidence interval calculated around the mean difference and plotted. Plots were then examined for any systematic difference. Regression analysis was then performed to see if there was a proportional bias across the range of means. Intraclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated for the two scales using a two factor mixed effects model and type consistency (McGraw & Wong, 1996; Shrout & Fleiss, 1979). ICC values less than 0.5 are indicative of poor reliability, values between 0.5
and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability (Koo & Li, 2016).

One aspect of validity was assessed using the known groups approach (DeVellis, 2003; Cappelleri, Zou, Bushmakin, Alvir, Alemayehu, & Symonds, 2013), separate analyses were carried out for gender, contact, dangerousness and reliability with total social acceptance score. In addition, for the social acceptance scale, we expected males, those with less prior contact, and those with higher expectations of dangerousness and unreliability would have higher levels of social distance. This was used this to help validate the social acceptance scale. To investigate differences between groups, t-tests were conducted to determine whether there were group differences in the predicted direction on the social acceptance scale. Effect size indices, which provide information about the degree of differences between the groups, were also calculated. For t-tests, a $d$ value suggests a small (.20 and below), medium (.30 to .70), or large (.80 and above) degree of difference (Cohen, 1988).

Once the scale items were selected, two confirmatory factor analyses were carried out on both the resulting 11-item stereotypic attributions scale and the 11-item social acceptance scale. Since items were ordered categorical variables, data were analyzed using diagonally weighted least squares (WLSMV), which is specifically designed for ordinal data. WLSMV makes no distributional assumptions about the observed variables (Brown, 2015). Model fit was determined based on statistical recommendations that Root Mean Square Error of Approximation (RMSEA) be below 0.08 for a reasonable fit (below 0.05 for a close fit), Comparative Fit Index (CFI) be equal to or greater than 0.95, and the standardized root mean square residual (SRMR), with a value of 0.05 or less being suggestive of an acceptable model (Bentler & Chou; Browne & Cudeck, 1993; Hu & Bentler, 1999). Although Chi-Square was reported, it is a problematic fit
index since it is sensitive to sample size and violations of normality (Joreskog, 1969). To obtain better fitting models secondary confirmatory factor analysis were conducted on both scales which allowed for correlated error variance to allow for methods effects.

3.4 Results

3.4.1 Sample Characteristics

Table 2 shows the characteristics of the sample. Females outnumbered males in all samples, but most in the confirmatory factor analysis. Grades ranged from 8-12 in both the exploratory and the confirmatory factor analysis but were limited to grades 11 and 12 in the test-retest analysis. Approximately 25% (23%-26%) responded they had a mental illness, and approximately one third (32%-35%) responded they did not or were unsure if they knew someone with a mental illness.

3.4.2 Item analysis

Descriptive statistics (see Appendix) were used to look at the distribution of scores; in particular, the frequency distributions, means, standard deviations, range, skewness, and kurtosis. No floor or ceiling effects were identified for either the Stereotypic Attribution items or the Social Acceptance Items. Stereotypic Attribution responses ranged across categories with a smaller proportion for the negative endorsements of agree or strongly agree. Almost all items showed a strong positive endorsement of either strongly disagree or disagree, with 9 of the 15 items showing over 50% positive endorsement. The highest positive endorsement was for the item “People with a mental illness could snap out of it if they wanted to” with a 79.6%
disagreeing or strongly disagreeing with the statement. Although somewhat high, the percent is
from a combination of two response options. There were few missing values, the highest being
1.3% for the questions “Most people with a mental illness are too disabled to work” and “You
can’t rely on someone with a mental illness”. Item means ranged from 1.85 to 3.31 and most
skewness and kurtosis values were within ± 1.

Social Acceptance responses ranged across categories with a smaller proportion for the
negative endorsements of agree or strongly agree. Almost all items showed a strong positive
endorsement of either strongly disagree or disagree, with 13 of the 17 items showing over 50%
positive endorsement. The highest positive endorsement was for the reversed coded item “I
would not mind it if someone with a mental illness lived next door to me” with 80.2% either
agreeing or strongly agreeing positive endorsement. There were few missing values. The highest
being 1.5% for the question “I would not mind it if someone with a mental illness lived next door
to me”. Means for the items ranged from 1.97 to 3.52 and most skewness and kurtosis values
were within ± 1.

3.4.3 Exploratory Factor Analysis

The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.84 for the Stereotypic
Attribution items and 0.88 for the Social Acceptance items indicating items were sufficiently
correlated and the sample was suitable for a factor analysis. Bartlett's test of sphericity was
significant $p<.001$. A parallel analysis confirmed a one factor solution was appropriate for both
scales. The magnitudes of factor loadings were examined, items with loadings less than .40 were
removed and the analysis was repeated. When loadings of less than 0.40 were excluded, 11 items
were retained for both Stereotypic Attribution the Social Acceptance scales. The final loadings
for the Stereotypic Attribution scale ranged from 0.48 to 0.67. Sixty one percent of the common variance was explained. For the Social acceptance scale loadings ranged from 0.57 to 0.76. Sixty seven percent of the common variance was explained.

The resulting the Stereotypic Attribution Scale (EFA2 on Table 3) contained four items pertaining to controllability of the illness, two referring to recovery, and 5 referring to violence and unpredictability. The resulting Social Acceptance Scale (EFA2 on Table 4) contained 7 items measuring desire for social distance and 4 measuring social responsibility for mental health issues. Total scores ranged from 11 (least stigmatizing) to 55, (most stigmatizing) for both scales. For the Stereotypic Attribution scale the mean total score was 25.3 (95% C.I. 24.9 – 25.6) (median 25.0), and for the Social Acceptance scale it was 24.4 (95% C.I. 24.0 – 24.7) (median 24.0).

Spearman’s rank order correlation was used to assess the association between the two scales. A statistically significant but moderately low correlation of 0.48 ($p<0.001$) indicated that the two scales share some common variance but measured somewhat different constructs.

3.4.4 Reliability Analysis

Internal consistency analyses were conducted using Cronbach’s alpha coefficient to obtain reliability estimates. Both scales performed well. For the Stereotypic Attribution Scale, Cronbach’s alpha was 0.80 which well above the minimum level of 0.60 of reliability for exploratory studies. Item rest correlations ranged from 0.39 to 0.55. Item rest correlations were all above 0.30 and below 0.80 so can be considered as making a meaningful non-redundant contribution to the scale. Alpha if item deleted values were all below the overall alpha suggesting that all items were contributing to the scale and that deleting any one of them would not increase
reliability. For the Social Acceptance Scale, Cronbach’s alpha was 0.85 which again is well above the minimum level of 0.60 of reliability for exploratory studies. Item rest correlations ranged from 0.46 to 0.63. Item rest correlations were all above 0.3 so again can be considered as making a meaningful contribution to the scale. Deleting any one item did not improve the overall alpha level again suggesting that all items were contributing to the scale.

3.4.5 Known Groups’ Differences Validation for the Social Acceptance Scale

The social acceptance subscale discriminated, as expected, between male and female students. Males had a significantly higher score ($p<.001$) (implying a greater desire for social distance). The effects size for gender was large (0.80). The scale also discriminated between knowing someone with a mental illness and, again as expected, those who know someone had significantly lower scores ($p<0.001$). The effect size for knowing someone was medium at 0.30. Having a close friend, having a family member or knowing someone else with a mental illness were all significant in their relationship to total scale score for social acceptance ($p \leq 0.05$). Effect sizes ranged from small (0.13) to medium (0.58) with the largest being for having a close friend, which may be the most intimate of the three. These findings give evidence for known groups validity.

People who believe individuals with a mental illness are dangerous or blameworthy tend to show more social distance. Effect sizes presented are for strongly disagreeing compared to each of the other agreement levels. For the first item, people with a mental illness are often more dangerous than the average person, effect sizes are moderate to large (range for 0.44 to 1.08) and increase in the predicted direction in a dose response fashion with those with the strongly agreeing having the largest mean value on the Social Acceptance Scale (28.1) and the largest
effect size (1.08). Similarly, Social Acceptance Scale scores differed in the predicted direction with those strongly agreeing with the second item, “You can’t rely on someone with a mental illness, having the highest mean score (32.3). Effect sizes were high ranging for 0.78 to 1.90, the largest for the comparison of the 2 extremes, strongly agree verses strongly disagree.

3.4.6 Test-Retest Reliability

Both scales preformed reliably. Individual items on each scale were summed to obtain a scale score. Mean scores on both scales did not significantly differ for either scale for the two administrations, correlations were moderate (0.73 and 0.82) and alpha’s were high (0.79 and 0.86). One sample $t$-tests were run to test whether the mean changes between scale scores were significant. The test was non-significant for both difference scores ($p=0.093$ and $p=0.065$) suggesting a certain level of agreement between the two measurements. Mean difference scores were plotted against average scale scores and 95% confidence intervals were calculated. Plots for both do not show a trend (which would be illustrated if more points were above or below the mean difference line), so there did not appear to be any systematic difference. Regression analysis was not significant for both ($t = 1.58; p=0.116$ and $t=1.423 \ p=0.157$) suggesting there was not a proportional bias across the range of means. The ICC for the mean scale score of the baseline compared to the follow up for the Stereotypic Attribution Scale was 0.75 (95% CI= 0.67 - 0.81) and for the Social Acceptance Scale it was 0.81 (95% CI=0.75 - 0.86), both indicating good reliability.

3.4.7 Confirmatory Factor Analysis
The first confirmatory factor analysis for the Stereotypic Attribution Scale produced the following results: Chi square model = 314, df = 44, \( p < .001 \); Root mean square error of approximation (RMSEA) =0.106; Comparative fit index (CFI) =0.91; Tucker Lewis Index (TLI) = 0.89; Standardized root mean squared residual (SRMR) = 0.065. Several of the key goodness of fit statistics indicated that the model was a poor fit for the data. The Chi-square statistic was significant (which often happens in samples over 200), RMSEA was well above the .08 threshold at 0.11, the CFI of was less than the desired 0.95 threshold, the TLI was less than 0.90 and the SRMR was greater than 0.05. On a positive side all items had correlations of greater than 0.40 (CFA1 on Table 3). To obtain a better fitting model, a second confirmatory factor analysis was conducted allowing for correlated error variances among 5 items referring to violence and unpredictability. Correlated errors were specified among these items’ residuals to reflect method effects caused by high content overlap. The results were as follows: Although the Chi square model =130, df = 34, \( p < .001 \) was still significant; RMSEA improved to 0.072; the CFI of 0.97 exceeded the desired threshold; the TLI of 0.95 exceeded the desired threshold; and the SRMR of 0.043 was less than 0.05 (CFA2 on Table 3). All of these together are suggestive of an acceptable model.

The first confirmatory factor analysis for the Social acceptance Scale produced the following results, Chi square model = 642, df = 44, \( p < .001 \); RMSEA = 0.15; CFI = 0.90; TLI = 0.95; the SRMR = .078. Again, several of the key goodness of fit statistics indicated that the model was a poor fit for the data. The Chi-square model = 86, df=29 was significant, RMSEA was well above the 0.08 threshold at 0.15, the CFI of was less that the desired 0.95 threshold and the SRMR was greater than 0.05. The TLI, on ther other hand, exceed 0.90, and all items had correlations of greater than 0.40 (CFA1 on Table 4). To obtain a better fitting model, a
second confirmatory factor analysis was conducted allowing for correlated error variances among 3 items that were reverse worded to reflect method effects. The results were as follows: Although the Chi square model = 86, df = 29, \( p < .001 \) was still significant; RMSEA improved to 0.06; the CFI of 0.99 exceeded the desired threshold; the TLI of 0.95 remained the same and exceeded the desired threshold; the SRMR of 0.024 was less than 0.05 (CFA2 on Table 4). All of these are together are suggestive of an acceptable model.

### 3.5 Discussion

Until this study, a scale had not been developed to assess the attitudes and intended behaviours of adolescents towards people with mental illnesses. We have developed and tested two scales in a broad sample of Canadian youth from different secondary schools across the country. We assessed validity using different approaches: face validation, content validation, criterion validation, discriminant validation, and convergent validation. In addition, the scales showed good internal consistency and, test-retest reliability in the samples studied. The development of the scales was guided by models of stigma arising from literature reviews focusing on how attitudes towards people with mental illness can be measured. We chose to adopt the measurement of clear outcomes, such as attitudes and social acceptance, which can be measured using the various dimensions of stigma.

The confirmatory factor analysis showed method effects corresponding to the reverse coded items on the social acceptance scale. In future use of the scale consideration may be given to rewording these items so they are all in the same direction (so that no reverse coding is needed).
A strength of our approach was that extensive stakeholder consultations were held with regards to the items to be included in the scales and items included two broad measurement domains of interest, stereotypic attributions (controllability of the illness, potential for recovery, and potential for violence and unpredictability) and social distance (desire for social distance and feelings of social responsibility for mental health issues); all of which were the key foci for the educational programs provided. As there is no gold standard for assessing attitudes of Canadian youth towards people with mental illnesses, these scales provide a good starting point for further testing and development as well as a pilot scale that can be used in the evaluation of programs aimed at reducing the stigma of mental illnesses in youth.

There is little evidence supporting ‘best practices’ in the field of anti-stigma programming. While there is a rich theoretical literature dealing with the stigmatization of mental illnesses, the published literature contains little in the way of ‘how to’ guidance. Few programs have been rigorously evaluated for their stigma-related outcomes. More often they have been assessed for customer satisfaction. Therefore, current evidence is insufficient to establish best practices, though a number of promising practices have emerged (Stuart, Arboleda-Florez, & Sartorius, 2012). Working in partnership with the Opening Minds anti-stigma program, these scales can assist existing programs to make better use of their evaluation data. Currently, much data are collected, but they are seldom analyzed or formally reported. By conducting an analysis of data that have been standardized across programs, using instruments with known psychometric properties, program staff gained a better understanding of the effectiveness of their interventions and the Mental Health Commission of Canada gained a better understanding of the effectiveness of contact-based approaches in stigma reduction as they are applied across Canada. By systematically evaluating programs that were ongoing in the field,
*Opening Minds* has contributed to much needed practical knowledge concerning what works, when, and where. If these approaches are demonstrated to be effective, then the Commission can work with programs to develop standardized teaching manuals and toolkits that could be disseminated widely as part of a nation-wide anti-stigma effort.

Addressing stigmatization in youth is also important in terms of building for the future. At least theoretically, it will have future benefits, perhaps addressing stigmatization in adult roles. Not only is adolescence a critical time for the emergence of mental illnesses, it is also important for the consolidation of stigmatizing attitudes. Implementation of anti-stigma education in high schools can promote inclusion in the school environment and reduce future stigmatizing behaviours. Accessing large numbers of young people through schools is efficient as the public health infrastructure to provide health promotion and prevention programs already exists. Anti-stigma programs can build on existing curricula or school activities. Schools also have the ability to offer a broad scope of companion interventions (Stuart, Arboleda-Florez, & Sartorius, 2012). Currently, school health and wellness programs do not include anti-stigma interventions, so they are a strategic, but under-used resource. If identifying programs that are effective in reducing stigma in students is possible, then it may also be possible for these to be included as part of the regular health curriculum and programming offered within Canadian school systems.

Future research with the scales should include its use in evaluating future mental illness anti-stigma related educational interventions. Variation in outcomes across programs could be explored to help understand the active key ingredients necessary for success in stigma reduction. Once key ingredients are identified they could be used by programs to improve their interventions. To support the validity of inferences about change and group differences in the
evaluation of contact-based interventions, further research should explore the measurement invariance in the context of a longitudinal factor model.

3.6 Conclusions

In summary, within the samples studied, the reliability of the scales assessing stereotypic attributions and social acceptance was established using Cronbach’s alpha for internal consistency as well as test-retest reliability showing stability of the scales over time. Validity was established using a number of methods. Face and content validity of the scales was established through literature review and expert input. Exploratory and confirmatory factor analysis as well as the exploration of known groups’ differences contributed to the assessment of the extent to which the scales measured the intended constructs and whether they correlated with other known criteria.

The scales provide a good basis for further validation as well as tools that could be used in the evaluation of programs aimed at reducing mental illness related stigma in adolescents. Given that there is growing recognition of the need to reduce the stigma towards those with a mental illness; these tools can be used by programs to evaluate their effectiveness in increasing positive attitudes and social acceptance of the mentally ill. Through the use of these tools, programs will be able to further develop and tailor their interventions to better suit students’ needs.

Declarations
Acknowledgements
On behalf of the Opening Minds Initiative of the Mental Health Commission of Canada, we would like to thank the schools, teachers, staff, students, community professionals and speakers
who participated in this project. This project was funded by the *Opening Minds* Anti-stigma Initiative of the Mental Health Commission of Canada, which is supported by a grant from Health Canada.

**Competing interests**
MK is funded through the *Opening Minds* anti-stigma initiative as a Research Associate and HS is the Senior Consultant to the *Opening Minds* Anti-stigma Initiative and partially funded through the Bell Canada Mental Health and Anti-stigma Research Chair.
3.7 References


Table 2 Characteristics for the Exploratory, Confirmatory and the Test-Retest samples

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>EFA</th>
<th>CFA</th>
<th>Test-retest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (N=1352)</td>
<td>% (N=576)</td>
<td>% (N=190)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Male</td>
<td>38.9% (521)</td>
<td>28.7% (68)</td>
<td>42.2% (78)</td>
</tr>
<tr>
<td>• Female</td>
<td>61.1% (817)</td>
<td>71.3% (169)</td>
<td>57.8% (107)</td>
</tr>
<tr>
<td>• Missing</td>
<td>14</td>
<td>339**</td>
<td>5</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 8</td>
<td>4.8% (64)</td>
<td>20.4% (103)</td>
<td></td>
</tr>
<tr>
<td>• 9</td>
<td>9.5% (127)</td>
<td>33.9% (171)</td>
<td></td>
</tr>
<tr>
<td>• 10</td>
<td>3.7% (49)</td>
<td>25.2% (127)</td>
<td></td>
</tr>
<tr>
<td>• 11</td>
<td>52.7% (713)</td>
<td>11.1% (56)</td>
<td>99.5% (189)</td>
</tr>
<tr>
<td>• 12</td>
<td>28.4% (378)</td>
<td>9.3% (47)</td>
<td>0.5% (1)</td>
</tr>
<tr>
<td>• Missing</td>
<td>21</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• 12</td>
<td>0.1% (1)</td>
<td>0.4% (1)</td>
<td></td>
</tr>
<tr>
<td>• 13</td>
<td>2.0% (27)</td>
<td>30.8% (72)</td>
<td></td>
</tr>
<tr>
<td>• 14</td>
<td>7.5% (100)</td>
<td>16.2% (38)</td>
<td></td>
</tr>
<tr>
<td>• 15</td>
<td>7.0% (93)</td>
<td>12.0% (28)</td>
<td></td>
</tr>
<tr>
<td>• 16</td>
<td>30.1% (402)</td>
<td>19.7% (46)</td>
<td>60.5% (115)</td>
</tr>
<tr>
<td>• 17</td>
<td>38.2% (510)</td>
<td>15.8% (37)</td>
<td>38.6% (73)</td>
</tr>
<tr>
<td>• 18+</td>
<td>15.1% (201)</td>
<td>5.1% (11)</td>
<td>0.5% (1)</td>
</tr>
<tr>
<td>• Missing</td>
<td>18</td>
<td>342***</td>
<td></td>
</tr>
<tr>
<td>Does someone you know have a mental illness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• I do</td>
<td>23.7% (313)</td>
<td>26.2% (149)</td>
<td>23.1% (43)</td>
</tr>
<tr>
<td>• Family member</td>
<td>24.5% (323)</td>
<td>17.6% (100)</td>
<td>13.4% (25)</td>
</tr>
<tr>
<td>• Close Friend</td>
<td>17.0% (225)</td>
<td>10.7% (61)</td>
<td>18.3% (34)</td>
</tr>
<tr>
<td>• Somebody else</td>
<td>21.8% (288)</td>
<td>21.4% (122)</td>
<td>22.6% (42)</td>
</tr>
<tr>
<td>• Uncertain</td>
<td>20.3% (268)</td>
<td>21.3% (121)</td>
<td>23.7% (44)</td>
</tr>
<tr>
<td>• No</td>
<td>11.4% (151)</td>
<td>13.0% (74)</td>
<td>10.8% (20)</td>
</tr>
<tr>
<td>• Missing</td>
<td>32</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

* Multiple responses accepted so items will not sum to 100%.

** one program did not ask gender for n=339

*** one program did not ask age n=339
Table 3  Stereotypic Attribution factor loadings from exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and re-specified confirmatory factor analysis (CFA2)

<table>
<thead>
<tr>
<th></th>
<th>EFA⁺</th>
<th>EFA²⁺</th>
<th>CFA¹⁺</th>
<th>CFA²⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most people with a mental illness are too disabled to work.</td>
<td>0.537</td>
<td>0.526</td>
<td>0.558</td>
<td>0.585</td>
</tr>
<tr>
<td>People with a mental illness tend to bring it on themselves.</td>
<td>0.619</td>
<td>0.624</td>
<td>0.586</td>
<td>0.617</td>
</tr>
<tr>
<td>People with mental illnesses often don’t try hard enough to get better.</td>
<td>0.653</td>
<td>0.637</td>
<td>0.664</td>
<td>0.702</td>
</tr>
<tr>
<td>People with a mental illness could snap out of it if they wanted to.</td>
<td>0.552</td>
<td>0.543</td>
<td>0.570</td>
<td>0.602</td>
</tr>
<tr>
<td>People with a mental illness are often more dangerous than the average person.</td>
<td>0.660</td>
<td>0.665</td>
<td>0.633</td>
<td>0.496</td>
</tr>
<tr>
<td>People with a mental illness often become violent if not treated.</td>
<td>0.568</td>
<td>0.585</td>
<td>0.660</td>
<td>0.495</td>
</tr>
<tr>
<td>Most violent crimes are committed by people with a mental illness.</td>
<td>0.479</td>
<td>0.482</td>
<td>0.564</td>
<td>0.499</td>
</tr>
<tr>
<td>You can’t rely on someone with a mental illness.</td>
<td>0.657</td>
<td>0.668</td>
<td>0.654</td>
<td>0.634</td>
</tr>
<tr>
<td>You can never know what someone with a mental illness is doing.</td>
<td>0.493</td>
<td>0.506</td>
<td>0.539</td>
<td>0.436</td>
</tr>
<tr>
<td>Most people with a mental illness get what they deserve.</td>
<td>0.540</td>
<td>0.532</td>
<td>0.648</td>
<td>0.680</td>
</tr>
<tr>
<td>People with serious mental illnesses need to be locked away.</td>
<td>0.599</td>
<td>0.596</td>
<td>0.665</td>
<td>0.693</td>
</tr>
<tr>
<td>*There are few effective treatments for mental illnesses.</td>
<td>0.154</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Most people who have a mental illness recover.</td>
<td>0.185</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*People with a mental illness seldom pose a risk to the public.</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*People with a mental illness tend to be treated unfairly.</td>
<td>0.242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Items to be removed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Kaiser-Meyer-Olkin (KMO) = 0.84, Bartlett’s statistic = $\chi^2$ (105) = 3157.1, P<0.000, Percent of common variance explained= 52.7%, Reliability estimate = 0.90

2 Kaiser-Meyer-Olkin (KMO) = 0.84, Bartlett’s statistic = $\chi^2$ (55) = 2899.8, P<0.000, Percent of common variance explained= 61.2%, Reliability estimate = 0.90

3Chi square mode = 314, df = 44, p < .001; Root mean square error of approximation =0.106; Comparative fit index (CFI) =0.91; Tucker Lewis Index (TLI) = 0.89; Standardized root mean squared residual (SRMR) = 0.065.

4Chi square model= 129, df = 34, p < .001; Root mean square error of approximation =0.072; Comparative fit index (CFI) =0.97; Tucker Lewis Index (TLI) = 0.95; Standardized root mean squared residual (SRMR) = 0.043.
Table 4  Social Acceptance factor loading from exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and re-specified confirmatory factor analysis (CFA2)

<table>
<thead>
<tr>
<th>Item</th>
<th>EFA¹</th>
<th>EFA²</th>
<th>CFA¹</th>
<th>CFA²</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would be upset if someone with a mental illness always sat next to me in class</td>
<td>0.719</td>
<td>0.730</td>
<td>0.768</td>
<td>0.808</td>
</tr>
<tr>
<td>I would not be close friends with someone I knew had a mental illness</td>
<td>0.730</td>
<td>0.762</td>
<td>0.805</td>
<td>0.845</td>
</tr>
<tr>
<td>I would visit a classmate in hospital if they had a mental illness</td>
<td>0.571</td>
<td>0.569</td>
<td>0.652</td>
<td>0.532</td>
</tr>
<tr>
<td>I would try to avoid someone with a mental illness</td>
<td>0.726</td>
<td>0.727</td>
<td>0.793</td>
<td>0.838</td>
</tr>
<tr>
<td>I would not mind it if someone with a mental illness lived next door to me</td>
<td>0.611</td>
<td>0.609</td>
<td>0.585</td>
<td>0.559</td>
</tr>
<tr>
<td>-If I knew someone had a mental illness I would not date them</td>
<td>0.556</td>
<td>0.572</td>
<td>0.651</td>
<td>0.690</td>
</tr>
<tr>
<td>I would not want to be taught by a teacher who had been treated for a mental illness</td>
<td>0.595</td>
<td>0.588</td>
<td>0.598</td>
<td>0.635</td>
</tr>
<tr>
<td>I would tell a teacher if a student was being bullied because of their mental illness</td>
<td>0.615</td>
<td>0.591</td>
<td>0.603</td>
<td>0.400</td>
</tr>
<tr>
<td>I would stick up for someone who had a mental illness if they were being teased</td>
<td>0.623</td>
<td>0.633</td>
<td>0.709</td>
<td>0.530</td>
</tr>
<tr>
<td>I would tutor a classmate who got behind in their studies because of their mental illness</td>
<td>0.695</td>
<td>0.712</td>
<td>0.755</td>
<td>0.603</td>
</tr>
<tr>
<td>I would volunteer my time to work in a program for people with a mental illness</td>
<td>0.629</td>
<td>0.645</td>
<td>0.768</td>
<td>0.659</td>
</tr>
<tr>
<td>*I would be afraid of someone I knew with a mental illness if I met them on the street when I was alone</td>
<td>0.357</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*I would tell an adult (such as a parent or school counsellor) if I thought that a friend was getting a mental illness</td>
<td>0.331</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*I would tell my parents if I thought I was getting a mental illness</td>
<td>0.321</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*I would not want to go to a doctor if I thought I had a mental illness</td>
<td>0.296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*I would not want my teacher to know if I had a mental illness</td>
<td>0.218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*I would not want my classmates to know if I had a mental illness</td>
<td>0.177</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Items to be removed

1 Kaiser-Meyer-Olkin (KMO) = 0.86, Bartlett’s statistic = $\chi^2 (136) = 6141.9, P < 0.000$, Percent of common variance explained= 63.1%, Reliability estimate = 0.90

2 Kaiser-Meyer-Olkin (KMO) = 0.88, Bartlett’s statistic = $\chi^2 (55) = 4451.7, P<0.000$, Percent of common variance explained= 66.9%, Reliability estimate = 0.93

32Chi square model 642, df = 44, p < .001; Root mean square error of approximation = .15; Comparative fit index = .90; Tucker Lewis Index (TLI) = 0.95; Standardized root mean squared residual (SRMR)= .078;

32Chi square model 86, df = 29, p < .001; Root mean square error of approximation = .06; Comparative fit index = .99; Tucker Lewis Index (TLI) = 0.95; Standardized root mean squared residual (SRMR) = .024
### Initial Pool of Questions

**Knowledge/attitudes**

- One in every hundred people have schizophrenia.
- One in one hundred people will develop schizophrenia over the course of their lifetime.
- Schizophrenia is a rare disorder and my chances of getting it are very low.
- 1 in 4 people will develop a mental health problem over the course of their lives.
- Most people with a mental illness could snap out of it if they wanted to.
- People with schizophrenia are likely to be violent.
- Are dangerous to the public because of violent behavior.
- People with mental health problems are likely to be violent.
- People with schizophrenia are likely to be violent.
- People with schizophrenia are violent and unpredictable.
- People with mental illnesses tend to be dangerous and unpredictable.
- People with mental illness are, by far, more dangerous than the general population.
- People with mental illnesses tend to be dangerous and unpredictable.
- There are few effective treatments available for the mentally ill.
- There are effective treatments for schizophrenia.
- Can be successfully treated without drugs using psychotherapy or social interventions.
- There are few effective treatments available for the mentally ill. Most people with a serious mental illness, can with treatment, get well and return to productive lives.
- People with schizophrenia cannot work.
- Can work in regular jobs.
- People with schizophrenia cannot work.
- People who are mentally ill are too disabled to work.
- People with mental illnesses are untrustworthy.
- People with mental illnesses are often treated unfairly.
- People with schizophrenia are often treated unfairly or stigmatized.
- People with schizophrenia have a split personality.
- Suffer from split or multiple personalities.
- People with schizophrenia have a split personality.
- Schizophrenia is the same as split or multiple personality disorder.
- Schizophrenia is a brain disease.
- Mental illness is an illness like any other.
- Street drugs and alcohol cause schizophrenia.
- Schizophrenia is a treatable lifelong disorder.
- Schizophrenia is caused by stress.
- People with schizophrenia usually need medication.
- Need prescription drugs to control their symptoms.
- People with schizophrenia usually need medication.
- Improved medications for people living with schizophrenia cures the disorder.
- People with schizophrenia are often mentally retarded or of lower intelligence.
- Tend to be mentally retarded or of lower intelligence.
- Schizophrenia is caused by poor parenting.
- People with schizophrenia should delay seeking medical treatment because the symptoms often go away on their own.
- Hear voices telling them what to do.
• Are a public nuisance due to panhandling, poor hygiene or odd behavior.
• Can be seen talking to themselves or shouting in city streets.
• People living with schizophrenia are at low risk of committing suicide.

**Recovery**

• Those who are mentally ill can learn to manage their illness.
• The key to managing mental illness is to seek help from medical professionals.
• In most cases, keeping up a normal life in the community helps a person with mental illness get better.
• Stigma associated with mental illness can slow down the recovery process.
• People who recover from mental illness were not really mentally ill in the first place.
• Recovery from mental illness is not realistic for many.

**Prejudice stereotypes/ attributions/ devaluation**

• Would you be afraid to talk to someone with mental health problems?
  • I would be afraid to approach someone if I knew that they had schizophrenia.
  • Would you feel afraid to have a conversation with someone who has schizophrenia?
  • Would you be afraid to talk to someone who had schizophrenia?

• Would you be upset or disturbed to be in the same class as someone with mental health problems?
  • Would you be upset or disturbed about working on the same job with someone who has schizophrenia?
  • Would you be upset or disturbed to be in the same class with someone who has schizophrenia?
  • I would be uncomfortable to find out that someone in my class had schizophrenia.
  • I would be upset if someone with a mental illness sat next to me in class.

• Would you be able to be friends with someone with mental health problems?
  • I would make close friends with someone who had a mental illness.
  • Would you be able to maintain a friendship with someone who has schizophrenia?
  • Would you make friends with someone who had schizophrenia?

• I would not go to a physician if I knew they had been treated for a mental illness.
• I would let someone with a mental illness baby-sit my children.
• I would not want someone with a mental illness to be a school teacher.
• I would go to a doctor if I thought I had a mental illness.
• I would not mind if someone with a mental illness lived next door to me.
• Would you marry someone with schizophrenia?

**Engagement – Social responsibility**

• People with mental illness experience significant prejudice in Canadian society
• Most Canadians need to adopt a more tolerant attitude towards people with mental illness.
• I would make a one-time donation to a charity to support mentally ill people.
• I would make a regular donation to a charity to support mentally ill people.
• I would sign a petition to support better programs for the mentally ill.
• I would support spending more tax dollars to improve services for the mentally ill.
• I would volunteer my time to work in an agency for the mentally ill.
• I would join an advocacy program to improve the rights of the mentally ill.
• I don’t think that Canada needs to allocate more resources towards supporting the mentally ill.

**Help Seeking**

Would you tell an adult if someone you knew started to have symptoms of schizophrenia?

Would you tell an adult if you know someone with a mental illness was being bullied?
<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you feel embarrassed or ashamed if your friends knew that someone in your family had schizophrenia?</td>
<td></td>
</tr>
<tr>
<td>Would you be embarrassed if your friends knew that a member of your close family had mental health problems?</td>
<td></td>
</tr>
<tr>
<td>Would you feel ashamed if people knew someone in your family has been diagnosed with schizophrenia?</td>
<td></td>
</tr>
<tr>
<td>I would feel uneasy going to a mental health practitioner (e.g., psychologists, psychiatrist, social workers etc.) because of what some people would think.</td>
<td></td>
</tr>
<tr>
<td>If I thought I needed mental health services, I would get it no matter who knew about it.</td>
<td></td>
</tr>
<tr>
<td>If you had a serious emotional problem, would you:</td>
<td></td>
</tr>
<tr>
<td>□ Definitely go for professional help</td>
<td></td>
</tr>
<tr>
<td>□ Probably go for professional help</td>
<td></td>
</tr>
<tr>
<td>□ Probably not go for professional help</td>
<td></td>
</tr>
<tr>
<td>□ Definitely not go for professional help</td>
<td></td>
</tr>
<tr>
<td>How comfortable would you feel talking about personal problems with a professional? MHS</td>
<td></td>
</tr>
<tr>
<td>□ Very comfortable</td>
<td></td>
</tr>
<tr>
<td>□ Somewhat comfortable</td>
<td></td>
</tr>
<tr>
<td>□ Not very comfortable</td>
<td></td>
</tr>
<tr>
<td>□ Not at all comfortable</td>
<td></td>
</tr>
<tr>
<td>How embarrassed would you be if your friends knew you were getting professional help for an emotional problem? MHS</td>
<td></td>
</tr>
<tr>
<td>□ Very embarrassed</td>
<td></td>
</tr>
<tr>
<td>□ Somewhat embarrassed</td>
<td></td>
</tr>
<tr>
<td>□ Not very embarrassed</td>
<td></td>
</tr>
<tr>
<td>□ Not at all embarrassed</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix

### Descriptive Statistics for Stereotypic Attribution Items

<table>
<thead>
<tr>
<th>Stereotypic Attribution items</th>
<th>1 Strongly disagree % (n)</th>
<th>2 Disagree % (n)</th>
<th>3 Unsure % (n)</th>
<th>4 Agree % (n)</th>
<th>5 Strongly agree % (n)</th>
<th>Missing (n)</th>
<th>Mean</th>
<th>Skewness/Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 – There are few effective treatments for mental illnesses</td>
<td>7.0% (94)</td>
<td>28.3% (382)</td>
<td>34.6% (467)</td>
<td>26.9% (363)</td>
<td>3.3% (44)</td>
<td>(2)</td>
<td>2.91</td>
<td>-0.062 / -0.661</td>
</tr>
<tr>
<td>Q3 – Most people with a mental illness are too disabled to work</td>
<td>23.0% (307)</td>
<td>51.5% (688)</td>
<td>18.1% (242)</td>
<td>6.7% (90)</td>
<td>0.6% (8)</td>
<td>(17)</td>
<td>2.10</td>
<td>0.691 / 0.346</td>
</tr>
<tr>
<td>Q4 – People with a mental illness tend to bring it on themselves</td>
<td>32.8% (441)</td>
<td>34.7% (446)</td>
<td>19.6% (264)</td>
<td>10.9% (146)</td>
<td>2.0% (27)</td>
<td>(8)</td>
<td>2.15</td>
<td>0.666 / -0.370</td>
</tr>
<tr>
<td>Q5 – People with mental illnesses often don’t try hard enough to get better</td>
<td>26.5% (357)</td>
<td>44.6% (601)</td>
<td>17.5% (236)</td>
<td>9.5% (128)</td>
<td>2.0% (27)</td>
<td>(3)</td>
<td>2.16</td>
<td>0.769 / 0.123</td>
</tr>
<tr>
<td>Q6 – People with a mental illness could snap out of it if they wanted to</td>
<td>44.1% (596)</td>
<td>35.4% (476)</td>
<td>13.4% (181)</td>
<td>5.2% (70)</td>
<td>1.7% (23)</td>
<td>(6)</td>
<td>1.85</td>
<td>1.135 / 0.911</td>
</tr>
<tr>
<td>Q7 – People with a mental illness are often more dangerous than the average</td>
<td>13.8% (186)</td>
<td>34.0% (457)</td>
<td>27.8% (374)</td>
<td>20.6% (277)</td>
<td>3.7% (50)</td>
<td>(8)</td>
<td>2.66</td>
<td>0.199 / -0.752</td>
</tr>
<tr>
<td>Q8 – People with a mental illness often become violent if not treated</td>
<td>7.7% (104)</td>
<td>28.0% (377)</td>
<td>38.4% (517)</td>
<td>23.0% (309)</td>
<td>2.8% (38)</td>
<td>(7)</td>
<td>2.85</td>
<td>-0.039 / -0.514</td>
</tr>
<tr>
<td>Q10 – Most violent crimes are committed by people with a mental illness</td>
<td>19.6% (264)</td>
<td>35.9% (483)</td>
<td>27.9% (375)</td>
<td>14.3% (193)</td>
<td>2.3% (31)</td>
<td>(6)</td>
<td>2.44</td>
<td>0.337 / -0.588</td>
</tr>
<tr>
<td>Q11 – You can’t rely on someone with a mental illness</td>
<td>17.6% (235)</td>
<td>46.9% (626)</td>
<td>24.6% (329)</td>
<td>9.7% (129)</td>
<td>1.2% (16)</td>
<td>(17)</td>
<td>2.30</td>
<td>0.528 / -0.057</td>
</tr>
<tr>
<td>Q12 – You can never know what someone with a mental illness is going</td>
<td>5.3% (71)</td>
<td>22.6% (304)</td>
<td>35.4% (477)</td>
<td>32.7% (440)</td>
<td>4.1% (55)</td>
<td>(5)</td>
<td>3.08</td>
<td>-0.235 / -0.543</td>
</tr>
<tr>
<td>Q14 – Most people with a mental illness get what they deserve</td>
<td>39.4% (530)</td>
<td>35.7% (480)</td>
<td>21.3% (286)</td>
<td>2.5% (34)</td>
<td>1.1% (15)</td>
<td>(7)</td>
<td>1.90</td>
<td>0.780 / 0.275</td>
</tr>
<tr>
<td>Q15 People with serious mental illnesses need to be locked away</td>
<td>46.8% (633)</td>
<td>27.8% (376)</td>
<td>18.7% (253)</td>
<td>4.6% (63)</td>
<td>1.8% (24)</td>
<td>(4)</td>
<td>1.86</td>
<td>0.998 / 0.376</td>
</tr>
</tbody>
</table>

* Item not included in the final scale
### Descriptive Statistics for Reverse Scored Stereotypic Attribution Items

<table>
<thead>
<tr>
<th>Stereotyped Attribution items (Reverse scored)</th>
<th>1 Strongly agree % (n)</th>
<th>2 Agree % (n)</th>
<th>3 Unsure % (n)</th>
<th>4 Disagree % (n)</th>
<th>5 Strongly disagree % (n)</th>
<th>Missing (n)</th>
<th>Mean</th>
<th>Skewness/Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>RQ2- Most people who have a mental illness recover</em></td>
<td>0.7%(10)</td>
<td>17.9%(241)</td>
<td>35.8%(483)</td>
<td>40.6%(548)</td>
<td>5.0%(67)</td>
<td>(3)</td>
<td>3.31</td>
<td>-0.223 -0.580</td>
</tr>
<tr>
<td><em>RQ9 - People with a mental illness seldom pose a risk to the public</em></td>
<td>2.5%(33)</td>
<td>23.4%(313)</td>
<td>43.4%(580)</td>
<td>25.2%(337)</td>
<td>5.5%(74)</td>
<td>(15)</td>
<td>3.08</td>
<td>0.101 -0.319</td>
</tr>
<tr>
<td><em>RQ13 - People with a mental illness tend to be treated unfairly</em></td>
<td>23.6%(317)</td>
<td>48.8%(654)</td>
<td>16.9%(227)</td>
<td>7.8%(104)</td>
<td>2.9%(39)</td>
<td>(11)</td>
<td>2.18</td>
<td>0.908 0.622</td>
</tr>
</tbody>
</table>

* Item not included in the final scale
### Descriptive Statistics for Social Acceptance Items

<table>
<thead>
<tr>
<th>Social Acceptance items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Missing</th>
<th>Mean</th>
<th>Skewness/Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q18 - I would be upset if someone with a mental illness always sat next to me in class</td>
<td>31.1%</td>
<td>40.1%</td>
<td>18.6%</td>
<td>8.9%</td>
<td>1.3%</td>
<td>(7)</td>
<td>2.09</td>
<td>0.716/-0.094</td>
</tr>
<tr>
<td>Q19 - I would not be close friends with someone I knew had a mental illness</td>
<td>32.7%</td>
<td>42.5%</td>
<td>19.1%</td>
<td>4.8%</td>
<td>1.0%</td>
<td>(9)</td>
<td>1.99</td>
<td>0.755/0.310</td>
</tr>
<tr>
<td>Q21 - I would try to avoid someone with a mental illness</td>
<td>26.5%</td>
<td>44.5%</td>
<td>20.7%</td>
<td>7.2%</td>
<td>1.0%</td>
<td>(14)</td>
<td>2.12</td>
<td>0.652/0.063</td>
</tr>
<tr>
<td>*Q23 - I would be afraid of someone I knew with a mental illness if I met them on the street when I was alone</td>
<td>15.2%</td>
<td>34.0%</td>
<td>31.8%</td>
<td>15.0%</td>
<td>4.0%</td>
<td>(14)</td>
<td>2.58</td>
<td>0.294/-0.486</td>
</tr>
<tr>
<td>Q24 - If I knew someone had a mental illness I would not date them</td>
<td>11.0%</td>
<td>22.9%</td>
<td>41.7%</td>
<td>19.6%</td>
<td>4.9%</td>
<td>(13)</td>
<td>2.84</td>
<td>-0.036/-0.390</td>
</tr>
<tr>
<td>Q25 - I would not want to be taught by a teacher who had been treated for a mental illness</td>
<td>19.6%</td>
<td>45.3%</td>
<td>24.2%</td>
<td>8.1%</td>
<td>2.9%</td>
<td>(11)</td>
<td>2.29</td>
<td>0.692/0.269</td>
</tr>
<tr>
<td>*Q29 - I would not want to go to a doctor if I thought I had a mental illness</td>
<td>18.8%</td>
<td>35.7%</td>
<td>26.7%</td>
<td>15.1%</td>
<td>3.7%</td>
<td>(13)</td>
<td>2.49</td>
<td>0.380/-0.566</td>
</tr>
<tr>
<td>*Q30 - I would not want my teacher to know if I had a mental illness</td>
<td>9.6%</td>
<td>24.5%</td>
<td>31.0%</td>
<td>27.4%</td>
<td>7.5%</td>
<td>(15)</td>
<td>2.99</td>
<td>-0.069/-0.755</td>
</tr>
<tr>
<td>*Q31 - I would not want my classmates to know if I had a mental illness</td>
<td>4.6%</td>
<td>10.6%</td>
<td>29.6%</td>
<td>38.5%</td>
<td>16.8%</td>
<td>(16)</td>
<td>3.52</td>
<td>-0.513/0.063</td>
</tr>
</tbody>
</table>

* Item not included in the final scale
## Descriptive Statistics for Reverse Scored Social Acceptance Items

<table>
<thead>
<tr>
<th>Social Acceptance items (reverse scored)</th>
<th>1 Strongly disagree % (n)</th>
<th>2 Disagree % (n)</th>
<th>3 Unsure % (n)</th>
<th>4 Agree % (n)</th>
<th>5 Strongly agree % (n)</th>
<th>Missing (n)</th>
<th>Mean</th>
<th>Skewness/Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ22 - I would not mind it if someone with a mental illness lived next door to me</td>
<td>29.5%(393)</td>
<td>50.7%(675)</td>
<td>12.4%(165)</td>
<td>5.6%(74)</td>
<td>1.9%(25)</td>
<td>(20)</td>
<td>1.99</td>
<td>1.084 1.337</td>
</tr>
<tr>
<td>*RQ26 - I would tell an adult (such as a parent or school counsellor) if I thought that a friend was getting a mental illness</td>
<td>11.1%(149)</td>
<td>42.1%(564)</td>
<td>33.0%(442)</td>
<td>11.0%(147)</td>
<td>2.8%(37)</td>
<td>(13)</td>
<td>2.52</td>
<td>0.455 0.014</td>
</tr>
<tr>
<td>*RQ27 - I would tell my parents if I thought I was getting a mental illness</td>
<td>17.1%(230)</td>
<td>43.9%(590)</td>
<td>23.3%(313)</td>
<td>10.5%(141)</td>
<td>5.1%(69)</td>
<td>(9)</td>
<td>2.43</td>
<td>0.689 -0.009</td>
</tr>
<tr>
<td>RQ28 - I would tell a teacher if a student was being bullied because of their mental illness</td>
<td>31.9%(427)</td>
<td>40.6%(544)</td>
<td>19.5%(261)</td>
<td>5.8%(78)</td>
<td>2.2%(29)</td>
<td>(13)</td>
<td>2.06</td>
<td>0.842 0.149</td>
</tr>
<tr>
<td>RQ32 - I would stick up for someone who had a mental illness if they were being teased</td>
<td>32.2%(431)</td>
<td>45.6%(610)</td>
<td>17.4%(233)</td>
<td>2.5%(34)</td>
<td>2.2%(29)</td>
<td>(15)</td>
<td>1.96</td>
<td>1.018 1.384</td>
</tr>
<tr>
<td>RQ33 - I would tutor a classmate who got behind in their studies because of their mental illness</td>
<td>18.6%(250)</td>
<td>40.1%(538)</td>
<td>28.4%(382)</td>
<td>9.4%(126)</td>
<td>3.5%(47)</td>
<td>(9)</td>
<td>2.39</td>
<td>0.546 -0.047</td>
</tr>
<tr>
<td>RQ34 - I would volunteer my time to work in a program for people with a mental illness</td>
<td>16.3%(219)</td>
<td>32.1%(430)</td>
<td>35.7%(478)</td>
<td>11.7%(157)</td>
<td>4.2%(57)</td>
<td>(12)</td>
<td>3.33</td>
<td>0.306 -0.316</td>
</tr>
</tbody>
</table>

* Item not included in the final scale
Chapter 4

A Utilization Focused Evaluation of Youth Contact-Based Stigma Reduction Programs

4.1 Abstract

This practice note details a utilization focused evaluation (Patton M. Q., 2008) of contact based mental health stigma reduction programs aimed at Canadian youth. Use of results produced during an evaluation process is a concern for evaluators. A use-oriented approach aims to ensure that a number of conditions are met to optimize the use of knowledge produced during the process. The goal of this practice note is not to present the evaluation results as such, but rather to report experiences in applying a utilization focused evaluation. It presents an example of an evaluation focused on the users and intended use of the evaluation results. It describes our approach as well as some of the conditions of the transfer of knowledge favorable to the use of results by users. Some successes of the approach are highlighted.

Keywords: mental illness stigma; social distance; stereotyping; prejudice; adolescents; evaluation, partnership, knowledge translation, research collaboration

4.2 The Evaluation Context

In 2001, the World Health Organization declared stigma as the “single most important barrier to overcome.” to improve mental health service delivery worldwide
In 2007, the Canadian Government created the Mental Health Commission of Canada with a 10-year mandate to provide national leadership in mental health reform. One of the pillars of activity was to implement an anti-stigma program. *Opening Minds*, the anti-stigma initiative of the Mental Health Commission, recognizes that stigma is primarily a problem of behaviours resulting in the unfair and inequitable treatment of people with a mental illness and their family members. Thus, the goal of the initiative is to effect change that would result in improvements in the day-to-day lives of those living with a mental illness. To accomplish this, the *Opening Minds* initiative identifies, documents, and disseminates leading practices in stigma reduction using networks of existing programs as community leaders. This was based on the recognition that there was already important anti-stigma work ongoing throughout Canada. They were also interested in fostering a network of practice so that local programs that were demonstrated to be effective could develop wider reach and contribute to a national effort, and program leaders could act as mentors to newly developing programs.

*Opening Minds* adopted grass roots, community development philosophy designed to build partnerships with existing community anti-stigma initiatives. If a national anti-stigma effort were to be sustained beyond the Commission’s mandate (in 2017), existing programs would have to be the driving force of this effort. This stands in contrast to many of the international anti-stigma programs ongoing at the time that centered activities on large, expensive, centrally administered media campaigns. Previous work with the *Opening Minds* program had demonstrated that large media campaigns
were too expensive to sustain and failed to produce the desired outcomes in Canadian audiences.

The Board of the Mental Health Commission identified four groups for targeted anti-stigma activity: youth, media, health care providers, and workers. This paper reports on the process and results of the youth evaluation. The details of the evaluation process, described below, are presented elsewhere in detail (Stuart, et al., 2014; Stuart, et al., 2014b).

4.3 Methods

We performed a utilization focused evaluation (Patton M. Q., 2008), which contends that an evaluation should be judged by its utility to its intended users. Thus, our evaluation was planned and conducted in alignment with our partners in an effort to enhance utilization, tool performance and decision making.

4.4 Identification of Stakeholders

The Opening Minds team selected 20 youth programs (delivering 29 interventions) based on a Request for Interest and subsequent international review of submissions. All of the programs chosen for participation in the evaluation network had an element of contact-based education involving people with lived experience of a mental illness. Contact based interventions aim at reducing stigma using direct interactions with people who have experience a mental illness (Couture & Penn, 2003). Initial contact with selected programs was made via a telephone interview. Once programs were chosen, the key program representatives were interviewed by Opening Minds staff to
further determine suitability for the initiative and to discuss a potential partnership agreement. They were then asked if they would like to attend a face-to-face meeting to be held in Toronto to further discuss partnership details. Subsequently, written partnership agreements between *Opening Minds* and each individual program were signed.

As part of the partnership agreement, programs committed to using the evaluation to improve program effectiveness, inform decisions about future programming, and support their efforts to access funding. During the Toronto meeting, there had been considerable discussion concerning the potential risks and benefits to participating in an evaluation network. The main benefits would be a third-party evaluation of outcomes and a formal report describing these. Programs considered that such information would support their funding requests. They were also supportive of the goal of having the researchers write up the results for academic publication in order to contribute to best practice literature. A major concern was what to do in case of unsuccessful results. They considered that a weak or negative outcome could damage their community relationships and threaten future funding. Programs wanted to retain the right to remain anonymous and to block circulation of their individual reports if the evaluation showed them to be unsuccessful. This formed a key part of the partnership agreement and was written into the evaluation proposal that received ethics clearance (Appendix B). Programs also agreed to collect evaluation data using standardized methods and scales (to be developed by *Opening Minds* affiliated researchers) and to collect the necessary evaluation data. The majority of programs were collecting performance data prior to and following the intervention, with no comparison groups, which they found feasible. Instruments were
not standardized across programs, making comparisons impossible, nor were they psychometrically tested. In some cases, programs collected data but did not have the time or expertise to conduct statistical analyses. Though there was the sense that the interventions were working, there was little in the way of systematically collected or analyzed data to demonstrate this. Once standardized data were collected, they were sent to the Queen’s University research team, which was funded by Opening Minds, for data entry, analysis, and report generation.

4.5 Overview of the Evaluation Process

To more thoroughly understand the programs, a qualitative study was undertaken to build a program logic model (figure 3) to indicate how the various parts of the program interact to produce an effect (see page 28). The research team worked with 18 contact-based educational programs implemented across Canada that targeted high school students, all formally affiliated with the Opening Minds anti-stigma initiative. Twenty in-depth interviews with stakeholders (program coordinators, speakers with lived experiences, family members) were conducted along with field observations of seven programs, and an extensive review of program materials. Data analysis involved collecting critical ingredients into domains for conceptual clarity and logic model building. Finally, content validation of the program logic model through a stakeholders’ review was conducted. A more detailed description of the various components can be found elsewhere (Chen, Koller, Krupa, & Stuart, 2016).
The inputs refer to the resources invested in a program that are necessary to form a supporting infrastructure for program activities. These include having secure funding, a strong team, good relationships with schools, good speaker preparation and coordination of activities. Processes are the activities of the program in relation to the intervention delivered. Who delivers it, what they say and do, how information is provided, whether materials are also provided, where the intervention is provided (e.g. a classroom setting, assembly, or summit), and when. Finally, the outcomes define the expected program effects. In this logic model, the outcomes are based on a series of levels, ranging from the simplest (student reaction to the program) to the most difficult to achieve (social changes). This logic model was developed to help guide program activities along some logical course in order to maximize their chances of success.

Given the lack of psychometrically tested and standardized instruments, it was agreed by the network partners that the first step would be to develop instruments that would capture the key dimensions of stigma addressed by the programs in conducting
contact-based education. Program stakeholders played an integral part in the
development of the resulting scales, from defining the domains of interest, providing
questions they were currently using, refining the wording of questions, and holding a
focus group to review potential scales. Several programs also participated in a pilot test
by collecting and providing data for psychometric analysis. The results of these analyses
are reported elsewhere (see Chapter 3).

4.6 Study Design

We did not alter the program’s data collection protocols so surveys were
conducted before and after the intervention without a comparison group. Comparison
data could not be collected feasibly by most of the programs. Several programs targeted
entire schools, and those that targeted classrooms, found it impossible to gain entrance
into additional classes to collect comparison data. We matched the surveys using a
unique, anonymous identifier. The evaluation design was approved by Queen’s
University Health Sciences and Affiliated Hospitals Research Ethics Board, and where
required, by appropriate School Board ethics/research committees (Appendix B).

Data were collected by program partners from October 5, 2010 to December 14,
2012. All questionnaires were set to Queen’s University. Data were double entered and
when discrepancies were found, original surveys were reviewed, and corrections were
made. Reversed phrased items were reverse coded. Initial data analyses were conducted
to assess the quality of the data.
We developed two unidimensional scales. The Stereotypic Attribution Scale contained 11 items measuring stereotypic attitudes towards people with a mental illness. The Social Acceptance Scale contained 11 items addressing social distance and social responsibility. For the Stereotypic Attribution Scale, Cronbach’s alpha was 0.80 and for the Social Acceptance Scale, Cronbach’s alpha was 0.85 indicating high internal consistency. This initial development and refinement of the scales took approximately two years to complete.

4.7 Knowledge Exchange

A Webinar was held to review the pilot test results from the scale development exercise. Participants included one or two representatives from each program, members from the Mental Health Commission, members from a consumer group, and researchers. Results were explained and discussed. Input was provided on what would be useful to see in individual program reports. Overall it was decided that reports that could be used for various audiences, such as programs, funders and schools, and for different reasons, such as program improvement or to gain program funding would be most useful. The Mental Health Commission’s Opening Minds initiative also required a format they could use to justify and promote a national rollout of contact based anti-stigma programs.

One program was chosen to help develop the report format. The program was chosen based on the fact they had the infrastructure, funding, and human capacity to devote the time required. Many of the programs, although wanting to participate more, did not have the resources. A draft report of the program specific results was sent to the
program and a time was set for a teleconference to discuss it in detail. Three members of
the evaluation team, two members from the Mental Health Commission’s *Opening Minds*
initiative and two program representatives participated. Over two hours, the report was
reviewed in detail with considerable discussion regarding the program and *Opening
Mind’s* stakeholders’ interpretation of the meaning of the results.

The group described the initial report draft showing a standard difference of
means analysis as too technical and academic. More important than the aggregated scale
results, they preferred to have a detailed breakdown of tabular results for each individual
question showing totals and percents. In addition, there was much discussion on how the
overall change should be presented, given that mean differences were not considered to
be informative against the patterns observed in the item responses. The approach
ultimately agreed upon was to assess the proportion of students who achieved an A-grade
(80% correct or higher) on each scale before and after the interventions. Given an
educational model, it was considered that this could be easily explained and understood
and that it would resonate well with school board officials and educators. This was also
discussed at length with the full multi-disciplinary research team working with *Opening
Minds* (those working with the other target groups). The draft was revised based on the
initial group discussion and second discussion was held which lasted approximately one
hour to finalize the format and structure.

Using the feedback gained from the initial review, each program was provided
with a draft report containing a description of their evaluation findings. A teleconference
was scheduled with *Opening Minds* staff, the evaluation team, and the program staff to
discuss the results. The conclusions were deliberately left blank and we worked with
program staff to develop these. Program staff were asked to provide a description of their program and bullet points identifying what they thought the main take home messages of the evaluation were. Staff who wished to have their names on the report (who contributed content in this way) were offered this opportunity. Approximately one hour was spent reviewing each report with the associated program. Twenty-nine reports were completed in this way. Programs were given the option to have their reports posted on the Mental Health Commission’s website. Those who agreed, had their reports posted. Table 5 below summarizes some of the key knowledge translation and knowledge use activities.

**Table 5. Examples of Some of the Knowledge Translation Activates**

<table>
<thead>
<tr>
<th>Activity Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial planning meeting held in Toronto with <em>Opening minds</em> and one or two members for each of the programs.</td>
</tr>
<tr>
<td>Monthly planning and progress meetings with <em>Opening Minds</em> via conference calls.</td>
</tr>
<tr>
<td>Monthly meetings were held with programs via conference calls.</td>
</tr>
<tr>
<td>During the instrument development phase extensive interaction with programs via e-mail, one-on-one telephone calls and conference calls.</td>
</tr>
<tr>
<td>Webinar to review pilot test results from the scale development.</td>
</tr>
<tr>
<td>Conference calls and email exchange to develop the format for the individual program reports.</td>
</tr>
<tr>
<td>Conference call with each program to review draft reports and provide feedback.</td>
</tr>
<tr>
<td>Provided final reports to each program which included program input.</td>
</tr>
<tr>
<td>Based on the finding from the two programs discontinued their approach, one made major changes and the remainder made minor changes.</td>
</tr>
<tr>
<td><em>Opening Minds</em> began fostering a network of practice replicating promising programs in other communities.</td>
</tr>
<tr>
<td><em>Opening Minds</em> had presented some of the results at a number of conferences, such as the International Stigma Conference in Ottawa in 2012 and the International Stigma Conference in Copenhagen in 2017.</td>
</tr>
</tbody>
</table>
4.8 Evaluation Challenges and How they Were Addressed

1. Getting agreement from Programs to risk a negative evaluation

A major concern was what to do if the evaluation showed the program to be unsuccessful. They considered that a weak or negative outcome could damage their community relationships and threaten future funding. Although we used all of the data that were collected, programs retained the right to remain anonymous and to not have their individual report circulated if the evaluation showed them to be unsuccessful. This formed a key part of the partnership agreement and was written into the evaluation proposal. In these cases, individual reports were not posted on the web site of the Mental Health Commission of Canada.

2. Getting ethics reviews from schools (passive vs active ethics)

One of the biggest challenges faced by our pilot project team was accessing schools for the purposes of evaluation. Most school Boards have Research Committees that must review and approve all research and evaluation undertaken within the school district. Often the teachers didn’t know the Research Committee contact person in their District. In addition, each school district had a different set of forms and requirements, requiring multiple iterations of documentation to be prepared. One of the key challenges
surrounded consent. Many schools require an active consent in which a letter is sent home with the student explaining the purpose of the intervention and the evaluation. The parent must sign the letter and return it in order for their child to be included. When we used active consent, we typically obtained less than a 20% approval. Because all of the children were eligible to receive the intervention, a 20% or less approval made the collection of evaluation data logistically impossible. There was no reliable way for program staff to identify the children who were supposed to be in the evaluation versus those who were not. Therefore, schools that required active consent could not be included in the evaluation. Many schools allowed for passive consent. ‘Passive’ consent occurs when a letter of explanation is sent home to the parents. The parent must sign and return a form if they do not want their child to be included in the evaluation. The passive consent procedure that was used did not result in a single parental refusal.

3. Matching surveys

It was necessary to match the pre-test survey to the posttest survey for every child included in the evaluation in order to conduct the appropriate matched analysis. This was logistically difficult because the researchers could not know the identity of the children. Therefore, we could not ask that they indicate their name on the form. We used a unique anonymous identifier composed of the initials of the first and last name, the gender, and the date of birth. Some school research committees objected to the date of birth (considering it a personal identifier) so in these cases we had to request the student’s age. In order to get a perfect match, the child had to fill out this portion of the survey correctly each time. Occasionally a teacher would agree to maintain a master list and link
the surveys, but this was rare. We found the matching process to be highly laborious and, in some programs, many of the surveys could not be matched. When a large proportion of data would have been lost because surveys could not be matched, we conducted an unmatched analysis for some programs, which may have the effect of decreasing statistical significance.

4. Time consuming to develop knowledge exchange strategy

A great deal of time was spent initially to gain trust and buy in from the programs. This was necessary within our utilization focused framework. Considerable time was spent with programs and Opening Minds in all aspect of this evaluation, from developing the scales, developing a report format that was understandable and useful to program report generation and knowledge exchange.

4.9 Conclusions

Successful school-based anti-stigma strategies could substantially improve students’ health status, healthy life expectancy and quality of life (Stuart H., 2006). Contact-based education may be a promising strategy for reducing stigma in middle and high school students. Brief contact-based interventions have the potential to promote and consolidate positive attitudes, reduce social distance and promote a sense of social responsibility among high school students.

Using a utilization focused approach, it has been possible to build capacity to undertake program evaluation with a range of stakeholder groups that were interested but
that did not have the resources or expertise at their disposal. Programs have benefited from the reflective stance required by the evaluation enterprise. Based on their results, several programs have discontinued activities that were demonstrated to be ineffective and many more have modified their approach.

**Acknowledgements**

On behalf of the *Opening Minds* Initiative of the Mental Health Commission of Canada, we would like to thank the schools, teachers, staff, students, community professionals and speakers who participated in this project. This project was funded by the *Opening Minds* Anti-stigma Initiative of the Mental Health Commission of Canada, which is supported by a grant from Health Canada.

**Competing interests**

MK is funded through the *Opening Minds* anti-stigma initiative as a Research Associate and HS is the Senior Consultant to the *Opening Minds* Anti-Stigma Initiative and partially funded through the Bell Canada Mental Health and Anti-Stigma Research Chair.
4.10 References


Chapter 5

Manuscript – Reducing mental health stigma in middle and high school adolescents: A non-randomized comparison trial

Abstract

**Background:** Adolescents have become an important target for contact-based anti-stigma programs. One strategy for reducing mental illness-related stigma in students is to include contact-based sessions in their educational curricula. In such sessions students are able to interact socially with a person who has a mental illness. We sought to evaluate the effectiveness of this strategy in one program that delivered contact-based intervention using three different models.

**Methods:** The study was a non-randomized comparison trial. Results are based on one program which delivered three different models of contact-based interventions (in class; an assembly; and a summit) and collected data from a comparison group. One hundred and thirty-one classrooms within 35 Ontario middle and high schools participated. The primary outcome, change in social acceptance, was assessed using a scale developed for the Canadian Mental Health Commissions *Opening Minds* initiative.

**Results:** A significant reduction in stigma among adolescents was observed in association with the contact-based educational sessions when compared to the comparison group ($p < .001$) (mean change of 2.0 for both the in-class model and the assembly model and a mean change of 5.5 for
the summit model). The effect sizes (Cohen’s $d$) for the three models were 0.43, 0.60 and 1.05 respectively. Clustering of data by school and program was assessed but was not significant.

**Conclusions:** Our results add to a growing literature supporting the effectiveness of contact-based strategies for stigma reduction in adolescents. In this study, all three interventions were associated with positive outcomes, but the summit model achieved greater gain in social acceptance over either in class or the assembly format.

**Keywords:** Mental Health, Social Stigma; Social Distance; Stereotyping; Prejudice; Adolescents; Contact-Based Education
5.1 Introduction

This paper reports on an evaluation that we undertook as part of the activities of the Mental Health Commission of Canada’s Opening Minds Anti-stigma Initiative. The overall approach taken by Opening Minds has been described elsewhere (Stuart, et al., 2014a; Stuart, et al., 2014b). Briefly, the approach was to identify existing anti-stigma programs that used contact-based education provided by individuals who have experienced a mental illness, evaluate their results, and then replicate the most promising approaches across the country. The Youth Pilot Project Network included more than 25 programs that signed formal partnership agreements with Opening Minds to participate in a third-party evaluation of their activities. Most programs could not collect comparison data, complicating interpretations. This analysis was carried out to examine the variability in outcomes of one program which delivered three different types of contact-based education programs targeting middle and high school students in Ontario, and to identify intervention and student characteristics associated with successful outcomes. This was the only program that used multiple interventions and obtained comparison data.

5.2 Background

The onset of many of mental health disorders occurs during childhood or adolescence. Fifty percent of people with a mental illness will experience its onset by age 11, and 75% before the age of 25 (Kessler, Berglund, Demler, Jin, Merikangas, & Walters, 2005). Mental illnesses are considered to be the most prevalent source of disability in adolescents. Estimates suggest 14% to 25% of youth have some form of mental illness (Health Canada, 2002; Waddell, Offord, Shepherd, Hua, & McEwan, 2002; Waddell, Shepherd, Chen, & Boyle, 2013; Gore, et al., 2011).
In 2011, it was estimated that the number of Canadian youth living with a mental illness was over 1 million and that by 2041, there will be almost 1.2 million children and adolescents between the ages of 9 and 19 living with a mental illness in Canada (Smetanin, P; Stiff, D; Briante, C; Adair, C; Ahmad, S; Khan, K,; 2011). There is also some evidence that Canadian youth may experience higher levels of emotional distress than youth in other countries. In a multi-country study conducted by the World Health Organization, Canadian students were among the most likely to report feeling depressed for a week or more, with estimates ranging from a quarter to over one third, depending on age and gender (World Health Organization, 1996). Thirty-seven percent of high school students in one Southeastern Ontario school district reported multiple symptoms of emotional distress (such as depression or anxiety) and 62% reported multiple stressors from school, work, parents, and friends (Stuart H., 2006). The Centre for Addiction and Mental Health’s Ontario Student Drug Use and Health Survey (OSDUHS) found that about one-fifth (21.9%) of students in grades 7 through 12 reported visiting a professional about a mental health issue at least once in the past year. This was a significant increase from 2011 (15.1%) and 1999 (12.4%). The same study found that 15.3% of Ontario students rated their mental health as fair or poor. The percentage of students rating their mental health as fair/poor in 2013 (15.3%) did not significantly differ from 2011 (13.7%). However, the 2013 percentage was significantly higher than that found in 2007 (11.4%), the first year of monitoring (Boak, Hamilton, Adlaf, Beitchman, Wolfe, & Mann, 2014).

Despite the high prevalence of mental health disorders among adolescents, they are the least likely age group to seek help (Polanczyk, Salum, Sugaya, & Rohde, 2015). One of the most significant barriers to seeking help and continuing with treatment is thought to be the stigma associated with mental illnesses (Angermeyer & Dietrich, 2006; Boyd, Katz, Link, & Phelan,
Fear of stigmatization is one of the reasons reported by young people as affecting their willingness to seek help for a mental illness (Corrigan P., 2004; Chandra & Minkovitz, 2007; Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004; Bowers, Manion, Papadopoulos, & Gauvreau, 2013; Gulliver, Griffiths, & Christensen, 2010). Many adolescents worry about others viewing them negatively, and fear being judged as weak (Chandra & Minkovitz, 2007). In addition, youth who do receive treatment are more likely than older respondents to report being stigmatized. Research examining a national sample of Canadians reported youth (age 12 to 25) who received treatment for a mental health problem in the year prior to the survey were more likely than older respondents to report being stigmatized as a result of a current or past mental or emotional problem. These results support the need for anti-stigma programs that target youth (Stuart, et al., 2014a).

Although much research has been devoted to the development and origins of stigma and its consequences, less is known about what makes an anti-stigma program successful. Identification and evaluation of effective methods for intervention are important to the development of national programs to reduce stigma (Stuart, Arboleda-Florez, & Sartorius, 2012).

Work to reduce the stigmatization associated with mental illnesses has been carried out in many ways and the literature currently identifies six main strategies. These are education, protest, contact, legislative reform, advocacy and stigma self-management (Corrigan, et al., 2001; Arboleda-Flórez & Stuart, 2012). Education aims to replace myths and stereotypes about mental illness with correct information (Corrigan P. W., 2004). Protest uses strategies to demonstrate and challenge social injustices (Larson & Corrigan, 2008). Contact reduces stigma
using direct interactions with people who have experienced a mental illness. Couture and Penn (2003) describe contact as an attempt to dispel inaccurate and negative beliefs about mental illnesses by placing people in direct personal contact with the stigmatized group. Legislative reform prohibits discrimination, improves protections, and offers accommodations (in areas such as employment, education and housing) for people with a mental illness. Advocacy is used to ensure people with a mental illness actually realize the rights and freedoms offered by legislation. Stigma self-management programs and supports aim to empower people with a mental illness to overcome their illness identity and find new personal meaning and social roles with the recognition that recovery means more than symptom control and disease management.

A meta-analysis conducted by Corrigan and colleagues (2012), which included protest or social activism, education of the public, and contact with persons with mental illness, concluded that contact had a better effect on attitudes in adults, whereas education had a better effect on attitudes in adolescents. It was suggested that beliefs about mental illnesses might not be as firmly developed in adolescents as they are in adults, making adolescents more likely to be responsive than adults to education. However, the majority of existing evidence supports contact-based education as a promising strategy for improving knowledge, attitudes, and behavioural intent toward people with a mental illness (Rickwood D., Cavanagh, Curtis, & Sakrouge, 2004; Sakellari, Leino-Kilpi, & Kalokerinou-Anagnostopoulou, 2011; Stuart H., 2006; Wei, Hayden, Kutcher, Zygmunt, & McGrath, 2013; Kolodziej & Johnson, 1996; Pinfold, Stuart, Thornicroft, & Arboleda-Flórez, 2005). The lack of standard measures and the variety of outcomes assessed has limited comparability across studies. In addition, the nature and quality of the interventions delivered was unknown.
Schachter et al. (2008) reviewed of the literature on the effects of school-based interventions on mental health stigmatization. They concluded that while they did find evidence for the benefits of school base-interventions they could not determine what works best, for whom, and under what circumstances (e.g., setting), and therefore could not recommend any single school-based intervention or intervention type. They noted the inconsistencies in the approaches to investigating the effects of interventions as a barrier. A lack of comparability was observed for each study parameter, which included the definition of the objectives, interventions (e.g., activities, materials, content), controls, populations, implementation, evaluation, outcomes, (validated) instruments, analyses, and control for confounders.

One of the program partners in the Opening Minds initiative was delivering contact-based interventions using 3 different models. The interventions followed the same study design and used the same standardized instruments to assess outcomes, so results are directly comparable across the interventions. It was also the only program that was able to obtain comparison data. This afforded us the opportunity to investigate the differences in outcomes based on type of intervention.

5.3 Methods

5.3.1 Study Design

This was a non-randomized trial that involved middle and high school students from 131 classrooms within 35 schools in Ontario. As schools or individual teachers invited programs to deliver interventions, the data were naturally clustered. In other words, clusters were not randomly selected but were naturally occurring and researchers did not have control over the selection of clusters as part of the study design.
Students received one of three different contact-based educational programs or no program. Students were selected into the intervention based on their teacher’s wishes. The primary outcome measure was at the individual level; improvement in feelings of social acceptance defined as the degree to which students were willing to interact with a classmate who had a mental illness. Social Acceptance was assessed at two time points. All students in class on the day of the survey were eligible to participate in the evaluation. For the intervention groups, students received their baseline survey immediately prior to the intervention and follow-up surveys were collected up to two weeks post intervention. For the comparison group, there was a two-week interval between the administration of baseline and follow-up surveys. A unique identifier was used to match baseline and follow-up measures.

5.3.2 The Interventions

The contact-based education involved learning about mental illnesses from people with first-hand experience. Each of the contact-based sessions employed people with lived experience who were in a state of recovery. Students and school staff who participated in the program received current mental health and illness information and heard stories and interacted personally with people who had been affected directly with a mental illness. All speakers participated in consistent training where they were guided through a process of writing their stories of hope and recovery, practicing the delivery in front of program staff, shadowed experienced speakers in program projects, experienced the question period with the students, and, when ready, participated in one or more of the interventions. Speakers were active members of the
program. The program received ongoing support from 12 service providers, 4 school boards, 6
speakers, and an agency that dedicates staff time to lead and coordinate efforts.

Three types of contact-based intervention were offered:

**Five day in class model:** Teachers were given facilitation guides and student workbooks to
guide participants through mental health lessons. Program members delivered a consistent
introductory session for each class that was involved in the 5-day program. The students and
teaching staff then followed the lesson plans as they prepared to hear the recovery stories from
two people who had been affected by a mental illness, which occurred on the fourth day. The last
module of the program was a debrief session, followed by a discussion of community resources
and “where to go for help.” The duration was 10 hours over the five days.

**Assembly model:** The assembly model invited all students in the school to participate in a large
assembly involving a contact based educational session that included a presentation on mental
health and wellness; myths and stereotypes; prejudice and discrimination; facts and truths. This
was followed by personal stories of hope and recovery from two people who had been affected
by a mental illness. Students were then invited to engage in a Question and Answer session. The
assemblies lasted 90 minutes.

**Summit model:** Participating schools were invited to bring 4 student volunteers and one staff
member volunteer to a public teaching hospital providing a range of specialized assessment and
treatment services to those living with complex and serious mental illnesses. They learned about
mental illness and the associated stigma via presentations, breakout room experiential exercises,
and heard the personal stories from speakers with lived experience. They then made plans on how they would design and deliver mental health awareness activities to their respective schools. Participants were given materials to help them plan their awareness activities. The summits lasted 6 hours.

5.3.3 Measures:

The Opening Minds survey for adolescents (described in Chapter 3) was the instrument used. It requires participants to complete an 11-item scale intended to measure stereotypic attributions toward the mentally ill and an 11-item scale measuring social acceptance of people with a mental illness. In each scale, participants respond to statements using a 5-point Likert type scale from “strongly agree” to “strongly disagree”. Each item on the scales are scored from 1–5, such that scale scores range from 11 (minimal stigma) to 55 (maximum stigma). The survey contains several additional items, including demographic information and items asking whether participants had prior contact with someone with a mental illness. The scales were developed and validated in an earlier study (see Chapter 3). Validation procedures included the development, testing and refinement of an item-pool, obtaining feedback from experts and focus groups, and psychometric evaluation. The internal reliability of the Social Acceptance scale was good. The alphas for baseline and follow-up measures were 0.86 and 0.86 respectively.

5.3.4 Primary Outcome Measure

The most appropriate method to analyze baseline and follow-up data is highly debated (Bonate, 2000). For this analysis, the primary outcome of interest was ‘learning gain’ defined as the change in score on a measurement form baseline to follow up (Williams & Zimmerman,
The 11-item social acceptance scale from the *Opening Minds* survey for adolescents was used to assess social acceptance of students toward people with mental illness. The scale measures social distance (7 items) and feelings social responsibility (4 items). Expressed beliefs about one's behaviors in a variety of situations has been considered a proxy for discrimination, though it is imperfect as the correlation between perceptions of social distance and discriminatory behaviors in this area is not known (see Chapter 3). In this case, learning gain was measured by the individual changes in scores on the social acceptance scale from baseline to follow-up. A higher score indicates greater learning gain. A major advantage of this method is ease of interpretation of the transformed variable as either a net gain or net loss in the overall score (Bonate, 2000). The change scores were found to be approximately normally distributed.

### 5.3.5 Potential Confounders and Effect Modifiers:

In this analysis we assessed the following variables as confounders or effect modifiers: gender, grade, stereotypic attitudes and prior contact with someone who has had a mental illness because these variables have been identified as key covariates in the literature (Corrigan & Watson, 2007; Pinfield, Toulmin, Thornicrof, Huxley, Farmer, & Graham, 2003; Schulze, Richter-Werling, Matschinger, & Angermeyer, 2003; Corrigan P. et al., 2002). Unfortunately we were not able to collect data on other potential confounding variables such as socioeconomic status, culture, race, religion, nor were we able to assess social desirability. School administrators and teachers were reticent to allow some of these variables and wanted to avoid a longer survey. Although things such as socioeconomic status, culture, race, religion have been shown to be related to stereotypic attributions and social acceptance it is unclear how they are related to the change that may occur after a contact-based intervention.
5.3.6 Ethics

For all students a passive consent procedure was used where a letter of explanation was sent home to the parents. The parent needed to sign and return a form if they did not want their child to be included in the evaluation. The evaluation design was approved by Queen’s University Health Sciences and Affiliated Hospitals Research Ethics and where required, by appropriate School Board ethics/research committees. No parents refused to have their child included in the evaluation.

5.4 Data Management and Analysis

The sample size was not predetermined. All students in class on the days of the intervention were included in the analysis. Of the 5420 baseline and follow-up surveys collected, 4292 (79%) were successfully matched using unique identifiers to give a sample of 2146 matched pairs. There are several reasons why the matching was unsuccessful. Students may not have been present in class during either of the measurement times, they didn’t use the same unique identifier, they were excluded because they did not complete the full baseline or follow-up survey, or teachers failed to give the complete follow-up survey. Other programs in the Youth network that kept close track of student responses reported that virtually no students refused to complete either the baseline or the follow-up surveys. For this analysis, no imputation of scale scores was done and we did not carry forward baseline scores to the follow-up scores (as in an intention-to-treat analysis).
This analysis was restricted to students with change scores. Given that, of the matched sample, a change score could not be computed for 16% of the students due to missing values. Because of the potential for selection bias, we compared the characteristics of those who remained in the matched analysis with those who were eliminated using two-way frequency distributions of proportions and 95% confidence intervals for categorical variables, and a difference of means for continuous variables (baseline stereotype score and social acceptance score).

Students were not randomly assigned to intervention or comparison situations. Given the potential for baseline differences, we also compared characteristics of those in the intervention and comparison group using two-way frequency distributions of proportions and 95% confidence intervals for categorical variables, and a difference of means for continuous variables (baseline stereotype score and social acceptance score).

Figure 4 summarizes the data exclusions.

Figure 4 Data exclusion
Because the data were potentially clustered by school and classroom, an empty model was used to determine the intraclass correlation coefficient (ICC), which provides an estimate of the proportion of the variation in the study outcome that was due to differences between schools and classrooms. ICCs were run for school and classroom on the entire sample and the again on the sample excluding the comparison groups. Second, bivariate relationships were examined between the study outcome and each covariate. Covariates that were associated with the outcomes at a level of \( p < 0.20 \) were included in the model building process. Possible confounders were identified using the change-in-estimate method. First, a crude estimate of the effect between the exposure and outcome was calculated. Each covariate was then added to the model separately to obtain an adjusted estimate that represented the degree of association between the exposure and outcome that remained after the influence of the covariate was removed. A difference between the adjusted estimate and the crude estimate of 10% or more indicated that confounding may be present. Confounders were included in the final model to produce a more reliable indicator of the association between the exposure and outcome. Finally, the multivariate model building process began with the introduction of the covariates and proceeded using a backwards elimination approach. To test for effect modification, interaction terms were created (the cross-product of the potential effect modifier and the exposure) and added to the regression.

5.5 Results

5.5.1 Sample characteristics

Table 6 presents the descriptive statistics of the sample for those who remained in the analysis versus those excluded from the analysis due to the inability to match baseline and follow
up surveys. No significant differences were found for any of the characteristics. There were slightly more females than males in both samples and almost 50% were from middle school. One quarter identified themselves as having a mental illness and almost 40% did not know anyone or were uncertain if they knew someone with a mental illness.

Table 6 Comparison Between Those Included and Excluded from Analyses Due to Inability to Match Baseline and Follow-up Surveys

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Has Change Score % (95% CI)</th>
<th>N= 1845</th>
<th>Missing Change score % (95% CI)</th>
<th>N = 301</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>44.7% (42.4 - 46.7)</td>
<td>819</td>
<td>47.0% (41.4 - 52.6)</td>
<td>139</td>
</tr>
<tr>
<td>Female</td>
<td>55.3% (53.0 - 57.6)</td>
<td>1013</td>
<td>53.0% (47.4 - 58.6)</td>
<td>157</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>13</td>
<td>--</td>
<td>5</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>28.7% (26.6 – 30.8)</td>
<td>513</td>
<td>26.3% (21.3 – 31.3)</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>22.5% (20.6 – 24.4)</td>
<td>415</td>
<td>21.0% (16.4 – 25.6)</td>
<td>63</td>
</tr>
<tr>
<td>9</td>
<td>11.7% (10.2 – 13.2)</td>
<td>216</td>
<td>14.3% (10.3 – 18.2)</td>
<td>43</td>
</tr>
<tr>
<td>10</td>
<td>10.5% (9.1 – 11.9)</td>
<td>194</td>
<td>11.7% (8.1 – 15.3)</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>12.9% (11.4 – 14.4)</td>
<td>238</td>
<td>12.3% (8.6 – 16.0)</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>14.5% (12.9 – 16.1)</td>
<td>267</td>
<td>14.3% (10.3 – 18.3)</td>
<td>43</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>2</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.2% (-0.0 – 0.4)</td>
<td>4</td>
<td>0.7% (-0.2 – 1.64)</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>20.5% (18.7 - 22.3)</td>
<td>376</td>
<td>18.7% (14.3 – 23.1)</td>
<td>56</td>
</tr>
<tr>
<td>13</td>
<td>25.4% (23.4 – 27.4)</td>
<td>467</td>
<td>22.1% (17.4 – 26.8)</td>
<td>66</td>
</tr>
<tr>
<td>14</td>
<td>15.8% (14.1 – 17.5)</td>
<td>291</td>
<td>19.1% (14.7 – 23.5)</td>
<td>57</td>
</tr>
<tr>
<td>15</td>
<td>7.0% (5.8 – 8.2)</td>
<td>128</td>
<td>9.4% (6.1 – 12.7)</td>
<td>28</td>
</tr>
<tr>
<td>16</td>
<td>14.9% (13.3 – 16.5)</td>
<td>273</td>
<td>14.7% (10.7 – 18.7)</td>
<td>44</td>
</tr>
<tr>
<td>17</td>
<td>12.9% (11.4 – 14.4)</td>
<td>237</td>
<td>11.7% (8.1 – 15.3)</td>
<td>34</td>
</tr>
<tr>
<td>18</td>
<td>3.3% (2.5 – 4.1)</td>
<td>615</td>
<td>3.7% (1.6 – 5.8)</td>
<td>11</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>8</td>
<td>--</td>
<td>2</td>
</tr>
<tr>
<td>Does someone you know have a mental illness*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do</td>
<td>20.1% (18.2 – 21.8)</td>
<td>362</td>
<td>20.3% (15.5 – 24.8)</td>
<td>55</td>
</tr>
<tr>
<td>Family member/Close friend</td>
<td>17.4% (15.7 – 19.1)</td>
<td>312</td>
<td>18.1% (13.6 – 22.4)</td>
<td>49</td>
</tr>
<tr>
<td>Friend</td>
<td>12.3% (10.8 – 13.8)</td>
<td>221</td>
<td>8.9% (5.7 – 12.1)</td>
<td>24</td>
</tr>
<tr>
<td>Somebody else</td>
<td>19.4% (17.6 – 21.2)</td>
<td>348</td>
<td>18.8% (8.2 – 15.4)</td>
<td>51</td>
</tr>
<tr>
<td>Uncertain</td>
<td>25.9% (23.9 – 27.9)</td>
<td>465</td>
<td>25.8% (20.9 – 30.7)</td>
<td>70</td>
</tr>
<tr>
<td>No</td>
<td>15.2% (13.7 – 16.8)</td>
<td>274</td>
<td>18.8% (14.3 – 23.2)</td>
<td>51</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>47</td>
<td>--</td>
<td>30</td>
</tr>
</tbody>
</table>
Table 7 indicates those excluded from the analysis had a slightly higher score on the stereotypic attribution scale than those who remained in the analysis indicating a propensity for slightly more negative views. There were no significant differences between the groups on the baseline social acceptance scale.

Table 7 Mean baseline scale scores

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Has Change Score</th>
<th>N</th>
<th>Missing Change</th>
<th>N</th>
<th>Mean Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Score</td>
<td></td>
<td>Mean Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(standard deviation)</td>
<td></td>
<td>(standard deviation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>Stereotypic Attributions</td>
<td>26.0 (7.12)</td>
<td>1722</td>
<td>27.1 (5.78)</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Social Acceptance</td>
<td>24.0 (6.72)</td>
<td>1845</td>
<td>24.8 (7.31)</td>
<td>154</td>
</tr>
</tbody>
</table>

Table 8 presents the descriptive statistics for those in the intervention groups and those in the comparison group. The groups were similar in terms of gender and whether they knew someone with a mental illness. There were slightly more females than males in both samples, approximately one quarter identified themselves as having a mental illness and almost 40% did not know anyone or were uncertain if they knew someone with a mental illness. The two samples were different in terms of age and grade. The intervention group was spread across grades with the exception of grade ten, while the comparison group was less spread out with the majority in grade ten (59%).
Table 8 Comparison between Intervention and the Comparison group

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention % (95% CI)</th>
<th>N= 1785</th>
<th>Comparison % (95% CI)</th>
<th>N = 361</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>44.5% (42.2 – 46.8)</td>
<td>789</td>
<td>47.6% (42.4 – 52.8)</td>
<td>169</td>
</tr>
<tr>
<td>Female</td>
<td>55.5% (53.2 – 57.8)</td>
<td>984</td>
<td>52.4% (49.1 – 59.3)</td>
<td>186</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>12</td>
<td>--</td>
<td>6</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>28.3% (26.2 -30.4)</td>
<td>505</td>
<td>24.1% (17.7 – 28.5)</td>
<td>87</td>
</tr>
<tr>
<td>8</td>
<td>23.5% (21.5 – 25.5)</td>
<td>419</td>
<td>16.3% (12.4 – 20.1)</td>
<td>59</td>
</tr>
<tr>
<td>9</td>
<td>14.5% (12.0 – 16.1)</td>
<td>259</td>
<td>0.0 (0.0 – 0.0)</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.9% (0.46 – 1.33)</td>
<td>16</td>
<td>59.0% (53.9 – 64.1)</td>
<td>213</td>
</tr>
<tr>
<td>11</td>
<td>15.3% (13.6 – 17.0)</td>
<td>273</td>
<td>0.6% (-0.2 – 1.4)</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>17.4% (15.6 – 19.2)</td>
<td>310</td>
<td>0.0 (0.0 – 0.0)</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>3</td>
<td>--</td>
<td>0</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.3% (0.0 – 0.5)</td>
<td>6</td>
<td>0.0 (0.0 – 0.0)</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>22.1% (21.2 – 24.0)</td>
<td>392</td>
<td>11.1% (7.9 - 14.3)</td>
<td>40</td>
</tr>
<tr>
<td>13</td>
<td>26.2% (24.2 – 28.3)</td>
<td>465</td>
<td>18.9% (14.9 – 22.9)</td>
<td>68</td>
</tr>
<tr>
<td>14</td>
<td>17.6% (15.8 – 19.4)</td>
<td>312</td>
<td>10.0% (6.9 – 13.1)</td>
<td>36</td>
</tr>
<tr>
<td>15</td>
<td>1.4% (0.9 - 1.9)</td>
<td>24</td>
<td>36.8% (31.8 – 41.8)</td>
<td>132</td>
</tr>
<tr>
<td>16</td>
<td>13.2% (11.6 – 14.8)</td>
<td>235</td>
<td>22.8% (18.5- 27.1)</td>
<td>82</td>
</tr>
<tr>
<td>17</td>
<td>15.3% (13.6 – 17.0)</td>
<td>271</td>
<td>0.3% (-0.2 – 0.9)</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>4.1% (3.1 – 4.9)</td>
<td>72</td>
<td>0.0 (0.0 – 0.0)</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>8</td>
<td>--</td>
<td>2</td>
</tr>
<tr>
<td>Does someone you know have a mental illness**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do</td>
<td>20.0% (18.1 – 21.9)</td>
<td>343</td>
<td>21.1% (17.8 – 26.3)</td>
<td>74</td>
</tr>
<tr>
<td>Family member</td>
<td>18.3% (16.5 – 20.1)</td>
<td>315</td>
<td>13.1% (9.6 – 16.6)</td>
<td>46</td>
</tr>
<tr>
<td>Close Friend</td>
<td>11.6% (10.1 – 13.1)</td>
<td>200</td>
<td>12.8% (9.4 - 16.2)</td>
<td>45</td>
</tr>
<tr>
<td>Somebody else</td>
<td>18.8% (17.0 – 20.6)</td>
<td>323</td>
<td>21.7% (17.4 – 26.0)</td>
<td>76</td>
</tr>
<tr>
<td>Uncertain</td>
<td>26.4% (24.4 – 28.4)</td>
<td>454</td>
<td>23.1% (18.8 – 27.4)</td>
<td>81</td>
</tr>
<tr>
<td>No</td>
<td>15.4% (13.7 – 17.1)</td>
<td>264</td>
<td>17.4% (13.5 – 21.3)</td>
<td>61</td>
</tr>
<tr>
<td>Missing</td>
<td>--</td>
<td>67</td>
<td>--</td>
<td>10</td>
</tr>
</tbody>
</table>

** Multiple responses accepted so items will not sum to 100%.
Table 8 indicates those in the intervention groups had a slightly higher score on the stereotypic attribution scale than those in the comparison group, indicating more negative views. There were no significant differences between the groups on the baseline social acceptance scale.

Table 9 Mean baseline scale scores interventions vs comparison group

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Intervention Mean Score (standard deviation)</th>
<th>N</th>
<th>Comparison Mean Score (standard deviation)</th>
<th>N</th>
<th>Mean Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotypic Attributions</td>
<td>26.3 (6.47)</td>
<td>1640</td>
<td>25.2 (5.92)</td>
<td>323</td>
<td>-1.13 (-1.85 - -0.41)</td>
</tr>
<tr>
<td>Social Acceptance</td>
<td>24.0 (6.77)</td>
<td>1663</td>
<td>24.1 (6.78)</td>
<td>336</td>
<td>0.40 (-0.73 – 0.86)</td>
</tr>
</tbody>
</table>

Table 10 shows the ICC’s for the empty model for classes nested within schools. A mixed effects regression model was performed with students nested in classes nested in schools, but it did not significantly change the estimates for type of intervention, so for ease of presentation, the results of the linear regression are presented.

Table 10 Intraclass Correlations

<table>
<thead>
<tr>
<th>Level</th>
<th>ICC</th>
<th>Std. Err.</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Class</td>
<td>School</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 11 presents the bivariate linear relationship between change in social acceptance scores and type of intervention as well as other potential covariates in the model. All
Interventions showed a significant positive change in social acceptance when compared to the comparison group. Baseline stereotype attributions and baseline social acceptance scores were also significantly related to the behavioral change scores with higher negative attitudes at baseline having a greater change at follow-up. People who knew someone with a mental illness also had higher change scores at follow-up. Gender and having a mental illness were not significantly related to change in social acceptance score, but were below the $p<0.20$ threshold, so were included in the model building process. Grade and age were not significant and did not meet the $p<0.20$ threshold. They were tested as potential confounders and effect modifiers. All changes of effect on outcome measures were below 10%. When testing for effect modification interaction terms were created. None were significant so were excluded.
Table 11 Bivariate linear relationships

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bivariate Models</th>
<th></th>
<th></th>
<th></th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>95 % Confidence Interval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison Group (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Assembly</td>
<td>2.11</td>
<td>0.37</td>
<td>1.39</td>
<td>2.83</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>In Class</td>
<td>2.09</td>
<td>0.31</td>
<td>1.48</td>
<td>2.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Summit</td>
<td>4.25</td>
<td>0.59</td>
<td>3.19</td>
<td>5.40</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.38</td>
<td>0.23</td>
<td>-0.71</td>
<td>0.83</td>
<td>0.09</td>
</tr>
<tr>
<td>Grade</td>
<td>0.08</td>
<td>0.06</td>
<td>-0.05</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.06</td>
<td>-0.07</td>
<td>0.18</td>
<td>0.37</td>
</tr>
<tr>
<td>Prior Contact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertain/ No (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Family member/ Close friend/</td>
<td>0.58</td>
<td>0.24</td>
<td>0.12</td>
<td>1.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Has a mental illness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.55</td>
<td>0.29</td>
<td>-0.01</td>
<td>1.12</td>
<td>0.06</td>
</tr>
<tr>
<td>Pre-exposure Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline stereotypic attributions</td>
<td>0.10</td>
<td>0.18</td>
<td>0.07</td>
<td>0.14</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Baseline social acceptance</td>
<td>0.22</td>
<td>0.02</td>
<td>0.18</td>
<td>0.25</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

– Reference category.
Table 12 presents the final results from the linear regressions. The initial model was run with the variables gender, having had contact, having a mental illness, pretest stereotypic attributions, and social acceptance scores as the remaining variables did not meet the $p<.20$. Variables were removed one by one. The first variable to be removed was prior contact ($p=.685$), the next model removed the having a mental illness ($p=.597$). After controlling for baseline social acceptance score, interventions had significantly different change scores than the comparison group.
Table 12 Final Model Based on the Linear Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.590</td>
<td>0.57</td>
<td>-5.71</td>
<td>-3.40</td>
</tr>
<tr>
<td>Exposure</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Comparison Group (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Assembly</td>
<td>2.01</td>
<td>0.36</td>
<td>1.31</td>
<td>2.71</td>
</tr>
<tr>
<td>In Class</td>
<td>2.03</td>
<td>0.31</td>
<td>1.43</td>
<td>2.62</td>
</tr>
<tr>
<td>Summit</td>
<td>5.31</td>
<td>0.56</td>
<td>4.20</td>
<td>6.41</td>
</tr>
<tr>
<td>Gender</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Male (Reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Female</td>
<td>-1.43</td>
<td>0.23</td>
<td>-1.89</td>
<td>-0.97</td>
</tr>
<tr>
<td>Baseline stereotypic attributions</td>
<td>-1.43</td>
<td>0.23</td>
<td>-1.89</td>
<td>-0.97</td>
</tr>
<tr>
<td>Baseline social acceptance</td>
<td>-0.64</td>
<td>0.02</td>
<td>-0.14</td>
<td>-0.24</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.02</td>
<td>0.26</td>
<td>0.34</td>
</tr>
</tbody>
</table>
5.5.2 Effect Size

Effect sizes were estimated for each intervention based on the data provided in Table 11. A medium effect size of 0.43 (95% CI: 0.30-0.56) was associated with the in-class intervention (Cohen, 1988). Translated, this effect size indicates that approximately 67% of the students that participated in the in-class program had greater improvement in their social acceptance scores compared to the average student who did not participate in the sessions (McGough & Faraone, 2009). Similarly, a medium effect of 0.45 (95%CI: 0.29-0.60) was associated with those participating in the assembly. For the summit the impact was greater. The effect size was 1.01 (95% CI: 0.76-1.23), which is considered a large effect size and indicates 84% of those who received the intervention had a greater change in their social acceptance scores compared to the average score in the comparison group.

5.6 Discussion

Schools present ideal sites to implement anti-stigma programs. Successful anti-stigma strategies could substantially improve students’ health status, healthy life expectancy, and quality of life (Stuart H., 2006). We examined contact-based education delivered across three intervention types, and a comparison group using the same design and measurement tools. We believe this to be the first study to examine 3 different types of interventions delivered by the same program using the same methods and measures. Each of the interventions diminished reported stigma among adolescent students when compared to a group that did not receive an intervention.
The intervention that showed the most positive gain was the summit model. This may be because the participants in this intervention were volunteers and potentially more engaged and invested in learning about mental illnesses and ways to reduce the stigma associated with it.

In Schachter et al.’s (2008) review of the literature on the effects of contact-based interventions they noted the inconsistencies in approaches to investigating the effects of interventions as a barrier to determine what works best. They pointed to things such as the lack of comparability of study parameters such as comparable validated evaluation instruments, populations studied, analysis used, and lack of controls. A strength of this study was that we were able to compare three types of interventions using the same validated instrument within the adolescent school population and compare the results to a comparison group. Given that all three interventions were given by the same program, they all shared a number of characteristics often thought to vary across programs and impact on a program’s ability to obtain positive results. Things such as having good speakers and stories, speaker training, speaker support, facilitator style, program staff, and program funding were all constant across the three interventions making it more likely the differences seen were due to intervention type and not other program specific characteristics.

Our study had several limitations.

Non-Random allocation to intervention and comparison group:

Students were not randomly allocated to intervention and comparison conditions, as this proved to be unfeasible in the school settings under study. As a result, there is the
possibility of group differences as a result of residual confounding from unmeasured variables. We were not able to include other potential confounders such as socioeconomic status race, culture or religion. However, we did include the major potential confounders identified in the literature. Further, comparison of intervention and comparison groups showed they were similar with respect to gender and prior contact; but dissimilar with respect to age and grade. This is because the comparison subjects were accrued from a limited number of classrooms, whereas the interventions were delivered to a wider range of students. Age and grade were not significant in the final model so were not retained as potential confounders.

Blinding:

Students were not blind to the intervention or comparison conditions. Those in the intervention groups knew they were getting an intervention. Because our outcome measure was self-reported, it may have been affected by social desirability bias. Students may have responded more positively on the posttest to appear more accepting of individuals with a mental illness producing inflated change scores. In addition, this might have been different across the different interventions, in particular the summit compared to the other two interventions as students self-selected into this intervention. Inclusion of a social desirability scale would have been helpful to rule out this possibility, but it would have increased respondent burden and rendered data collection too burdensome. Thus, we are unable to rule out social desirability as a competing explanation for our findings.

Selection bias:
Matching surveys proved to be difficult. We were able to match 79%. Of those that we could not match 69% were baseline surveys while 31% were post intervention surveys. While we could not match 21% overall, comparisons of characteristics between matched and unmatched sample did not show any significant differences suggesting that associations were not distorted, and selection bias due to inability to match surveys was not a problem.

Misclassification:

Ideally, interventions would be assessed at the level of behaviour rather than attitudes, but this was not feasible in for this study. Thus, we cannot be sure that differences in self-reported perceptions of discrimination would correlate with actual behaviours. This is a difficulty endemic in this field of study and it does not preclude comparing standardized effect sizes with other research using similar self-perceived social acceptance measures. If misclassification did occur, it would not have differed across groups and would therefore be non-differential misclassification which could have the potential of biasing the estimate either towards or away form the null value.

Lack of Follow-up:

This study was short-term. Lack of follow-up did not allow us to determine the long-term effects of the interventions and whether long-term effects differed by type of intervention. Thus, we cannot comment on the need for or timing of ‘booster’ sessions.
Internal Validity and Generalizability:

Participants were not randomly allocated to the intervention or comparison group. As such, unmeasured demographics and other variables were not controlled, possibly affecting the internal validity of the present research. By including the untreated comparison group, threats such as maturation, history, and regression to the mean would occur in both groups, thus, in essence, balancing the groups in terms of these threats and therefore, observed group differences are more likely to represent treatment effects.

This evaluation was carried out for a specific program to investigate the different effects if the three different contact-based interventions they deliver. If the delivery of these interventions adheres to the protocol or program model they were using it is believed they would be generalizable to the youth population they target.

Although many contact-based interventions implement similar interventions, there may be considerable variation in various aspects of their interventions. Programs may differ in their infrastructure and how they deliver their interventions, which could have implications for generalizability. These may include such things as program stability, the training and support provided to program staff, the stories shared, their ability to model recovery, to engage their audience, and to manage active participation through questions and discussion. This variations may affect a programs ability to change reported social acceptance. To gain a better understanding of the important aspects that are associated with successful interventions future research should explored program variations to determine which ones are the most important in contributing to change.
5.7 Conclusion

Addressing stigmatization in youth is important in terms of building for the future. At least theoretically, it could have future benefits, perhaps addressing stigmatization in adult roles. Not only is adolescence a critical time for the emergence of mental illnesses, it is also important for the consolidation of stigmatizing attitudes. Implementation of anti-stigma education in high schools can promote inclusion in the school environment and reduce future stigmatizing behaviours. Accessing large numbers of young people through schools is efficient as the public health infrastructure to provide health promotion and prevention programs already exists. Anti-stigma programs can build on existing curricula or school activities. Schools also have the ability to offer a broad scope of companion interventions (Stuart, Arboleda-Florez, & Sartorius, 2012). Currently, these programs do not include anti-stigma interventions, so they are a strategic, but under-used resource. If identifying programs that are effective in reducing stigma in students is possible, then it may also be possible for these to be included as part of the regular health curriculum and programming offered within Canadian school systems. Future research should continue to investigate what makes contact programs work and for whom.

Overall, the study results are encouraging for the effectiveness of all three of the programs interventions at improving social acceptance of people with a mental illness. It is important to keep in mind that participant were volunteers. In the in-class and assembly models participation was determined by the school and the teacher. Although the summit model appeared to have the greater effect on increasing social acceptance, the
participants choose to be there and as such this model reached a smaller audience. The other two models have the potential to reach a greater number of people.

**Declarations**

**Acknowledgements**

On behalf of the *Opening Minds* Initiative of the Mental Health Commission of Canada, we would like to thank the schools, teachers, staff, students, community professionals and speakers who participated in this project. This project was funded by the *Opening Minds* Anti-stigma Initiative of the Mental Health Commission of Canada, which is supported by a grant from Health Canada.

**Competing interests**

MK is funded through the *Opening Minds* Anti-stigma initiative as a Research Associate and HS is the Senior Consultant to the *Opening Minds* Anti-stigma Initiative and partially funded through the Bell Canada Mental Health and Anti-stigma Research Chair.
5.8 References


Smetanin, P; Stiff, D; Briante, C; Adair, C; Ahmad, S; Khan, K.; (2011). *The life and economic impact of major mental illnesses in Canada: 2011 to 2041*. RiskAnalytica, on behalf of the Mental Health.


Discussion and Conclusions

This program of research was undertaken under the auspices of the *Opening Minds* anti-stigma initiative of the Mental Health Commission of Canada as part of its then, 10-year mandate to reduce stigma in Canada. The approach taken by the *Opening Minds* initiative incorporated a grassroots community development philosophy with clearly defined target groups, in this case adolescents, with contact-based education as the key intervention to reduce stigma (prejudice and discrimination). Contact based education programs share the goal of reducing mental health related stigma. These approaches center on people with lived experiences of a mental illness telling their personal recovery stories to students who then have an opportunity to ask questions and engage in active discussion. A complete description of this initiative is contained elsewhere (Stuart, et al., 2014a; Stuart, et al., Opening Minds in Canada: Targeting change, 2014b).

In the first phase of this research, two 11-point scales were developed, pilot tested, and preliminary psychometric testing was performed. One scale measured stereotypic attributions while the other measured social acceptance of those with a mental illness. The scales were pilot tested and revised through several iterations involving program staff. Both scales loaded heavily on a single factor and showed excellent internal consistency. Correlations examined those who identified people with a mental illness as dangerous (from the first sub-scale) with the total social distance score on the second sub-scale, hypothesizing that those with higher expectations of dangerousness
would also have higher levels of social distance. This was the case and helped to validate the social distance sub-scale. In addition, the social acceptance subscale discriminated, as expected, between male and female students. Males had a significantly higher score ($p < .001$) (implying a greater desire for social distance). The effect size for gender was large (0.80). The scale also discriminated between knowing someone with a mental illness and again, as expected, those who know someone had significantly lower scores ($p < 0.001$). The effect size for knowing someone was medium at 0.30. Having a close friend, having a family member or knowing someone else with a mental illness were all significant in their relationship to total scale score for social acceptance ($p \leq 0.05$). Effect sizes ranged from small (0.13) to medium. Cronbach’s alpha was used demonstrate the reliability of the two subscales in the questionnaire. For the Stereotypic Attribution Scale, Cronbach’s alpha was 0.80 which well above the minimum level of 0.60 of reliability for exploratory studies. Item rest correlations ranged from 0.39 to 0.55. Item rest correlations were all above 0.3 and below 0.8 so can be consider as making a meaningful non-redundant contribution to the scale. Alpha if item deleted values were all below the overall alpha suggesting that all items were contributing to the scale deleting, any one of them would not increase reliability. For the Social Acceptance Scale, Cronbach’s alpha was 0.85 which again was well above the minimum level of 0.60 of reliability for exploratory studies. Item rest correlations ranged from 0.46 to 0.63. Item rest correlations were all above 0.3 so again can be considered as making a meaningful contribution to the scale. Deleting any one item did not improve the overall alpha level again suggesting that all items are contributing to the scale.
The newly developed scales were used by the programs to conduct outcome evaluations of their activities. Data were forwarded to Queen’s University for analysis and report development. Each program received an individually tailored copy of their findings for each of the interventions evaluated. Program staff participated in an online discussion of the results and participated in their interpretation. Where program staff contributed heavily to the report, they were included as co-writers. An extensive knowledge exchange exercise was undertaken to ensure that the analysis, results, and reports were understandable to program staff and would meet their needs. Results were presented in non-academic language, in easy to interpret graphs and figures.

Overall improvements outweighed any negative results that the students’ may have had on both scales. In addition, the lower (more stigmatizing) pretest scores were associated with less change, and those with the very lowest scores were more likely to shift in a negative direction. In future work, it will be important to understand why some students react neutrally or negatively. We consider that it may have something to do with their readiness to learn (or readiness to change). Students who come to the class with more fixed stereotypes may need to be targeted earlier in their education, perhaps even as early as elementary school. In order to examine this concept further, we will need to develop and test a measure of readiness to change as we have not been able to find one that is appropriate in this context.

By involving programs in an evaluation network, they were encouraged to critically reflect on their evaluation data in a way that would strengthen their
interventions. Based on their findings, two programs discontinued their approach, another made important changes to the speaker selection and training, and the remaining programs deliberately tried to improve areas where they considered they were not achieving the desired level of improvement. Based on their experiences, program personnel have become leaders for their peers. This approach promotes sustainable development.

The Mental Health Commission Opining Minds Initiative began fostering networks of practice, so local programs that were demonstrated to be effective could develop a wider reach and contribute to a national effort and program leaders could act as mentors to newly developing programs. A number of programs showed positive results and *Opening Minds* has been working on replicating some of these programs in other communities. For example, *Opening Minds* has replicated the five class curriculum program in elementary and high schools in Yellowknife and Winnipeg. Major national summits have been held (entitled Headstrong Summits) to facilitate the upscaling and replication of this model. They are also working to incorporate the best elements from other contact-based education programs into their overall process model and toolkits.

There were several challenges in the conduct of this research that need to be considered. One of the first was getting agreement from programs to risk a negative evaluation. Programs considered that a weak or negative outcome could damage their community relationships and threaten future funding. They wanted to retain the right to remain anonymous and to block the circulation of their individual reports if the evaluation showed them to be unsuccessful.
Accessing schools for the purposes of evaluation was another. Many schools require an active consent. When we used active consent, we typically obtained less than a 20% approval. Because all the children were eligible to receive the intervention, a 20% or less approval made the collection of evaluation data logistically impossible. There was no reliable way for program staff to identify the children who were supposed to be in the evaluation verses those that were not. Therefore, schools that required active consent were eliminated from the evaluation.

We found the matching process to be highly laborious and, in some programs, many of the surveys could not be matched. We were required to use a unique anonymous identifier. In order to get a perfect match, the student had to fill out this portion of the survey correctly each time and many did not.

Another key challenge was there was no freedom to modify the intervention programs currently delivered or implemented. In addition, schools were reluctant to allow us to ask extensive questions, and School Boards were strict and protective of school time, allowing minimal data to be collected. It proved to be difficult to obtain comparison group data. This would have required teachers who were not receiving the program to allow program staff into their classrooms to collect surveys at two points in time. Programs were unable to obtain agreement from teachers to do this and most programs did not have the additional time to dedicate to collecting comparison group data. A non-randomized trial design was chosen. Although not without limitations, this approach provided insight into practical effectiveness, whether the interventions appeared to be working in everyday practice. Once this level of effectiveness has been established, the focus can shift towards fine-tuning the interventions and toward more rigorous study.
designs. As it is unlikely that stereotypes and feelings of social distance can spontaneously change on their own (stigma tends to be enduring), we can have some confidence that changes observed can be attributed to the program interventions.

When focusing on one program that delivered contact-based education using three different formats we found that all three showed an increase in social acceptance when compared to the comparison group, with the Summit model (involving student volunteers) showing the greatest learning gains.

6.1 Contributions

This thesis resulted in the creation of a short measurement instrument that captures the important dimensions of stigma addressed by program partners, stereotypic attributions towards and social acceptance of those with a mental illness. It is relatively short and feasible to use within a classroom setting to evaluate program effectiveness. All programs received program specific results and based on the findings from the two programs, discontinued their approach, one made major changes and the remainder made minor changes. The Mental Health Commission, through Opening Minds, now has evidence informed programs to scale up and they have begun fostering a network of practice replicating promising programs in other communities.
6.2 Future research directions

Validation of scales is an ongoing process. Further validation of measurement structure of the two scales should be carried out. In particular it will be important to examine the invariance of their structural properties over time and across different groups (such as gender).

Not all contact-based interventions had positive results. In the future it will be important to identify and validate program characteristics most predictive of positive outcomes.

We don’t know why some students change more than others (and some not at all). In future work, it will be important to isolate predictors of change so that programs may be more precisely targeted to students’ learning needs. We are considering the concept of ‘readiness to change’ in the context of anti-stigma programming as a potential predictor of outcomes. In order to examine this further, we will need to develop and test a measure of readiness to change (we have not been able to find one that is appropriate in this context).

6.3 Conclusions

In summary, within the samples, the reliability of the scales assessing stereotypic attributions and social acceptance was established using Cronbach alpha for the internal consistency of the scale as well as test retest reliability showing stability of the scales over time. Validity was established using a number of methods. Face and content validation of the scales was established through literature review and expert input.
Exploratory and confirmatory factor analysis as well as the exploration of known groups’ differences contributed to construct and criterion validation. Given that there is growing recognition of the need to reduce the stigma towards those with a mental illness, these tools can be used by programs to evaluate their effectiveness in increasing positive attitudes and social acceptance of the mentally ill, and further develop and tailor their interventions to better suit students’ needs.

This study also showed that using a utilization focused evaluation approach and involving key stakeholder in the process right from the beginning enhanced knowledge exchange. Finally, when focusing on one program that delivered contact-based education using three different formats we found that all three showed an increase in social acceptance when compared to the comparison group, with the Summit model (involving student volunteers) showing the greatest learning gains.
6.4 References


Smetanin, P; Stiff, D; Briante, C; Adair, C; Ahmad, S; Khan, K:. (2011). *The life and economic impact of major mental illnesses in Canada: 2011 to 2041*. RiskAnalytica, on behalf of the Mental Health.


Appendix A
Program Evaluation Standards


Utility Standards

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

- **U1 Evaluator Credibility** Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- **U2 Attention to Stakeholders** Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
- **U3 Negotiated Purposes** Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- **U4 Explicit Values** Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- **U5 Relevant Information** Evaluation information should serve the identified and emergent needs of stakeholders.
- **U6 Meaningful Processes and Products** Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.
- **U7 Timely and Appropriate Communicating and Reporting** Evaluations should attend to the continuing information needs of their multiple audiences.
- **U8 Concern for Consequences and Influence** Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Feasibility Standards

The feasibility standards are intended to increase evaluation effectiveness and efficiency.

- **F1 Project Management** Evaluations should use effective project management strategies.
- **F2 Practical Procedures** Evaluation procedures should be practical and responsive to the way the program operates.
- **F3 Contextual Viability** Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.
- **F4 Resource Use** Evaluations should use resources effectively and efficiently.

Propriety Standards

The propriety standards support what is proper, fair, legal, right and just in evaluations.

- **P1 Responsive and Inclusive Orientation** Evaluations should be responsive to stakeholders and their communities.
• **P2 Formal Agreements** Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.

• **P3 Human Rights and Respect** Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.

• **P4 Clarity and Fairness** Evaluations should be understandable and fair in addressing stakeholder needs and purposes.

• **P5 Transparency and Disclosure** Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.

• **P6 Conflicts of Interests** Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.

• **P7 Fiscal Responsibility** Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

**Accuracy Standards**

The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.

• **A1 Justified Conclusions and Decisions** Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.

• **A2 Valid Information** Evaluation information should serve the intended purposes and support valid interpretations.

• **A3 Reliable Information** Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.

• **A4 Explicit Program and Context Descriptions** Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.

• **A5 Information Management** Evaluations should employ systematic information collection, review, verification, and storage methods.

• **A6 Sound Designs and Analyses** Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.

• **A7 Explicit Evaluation Reasoning** Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.

• **A8 Communication and Reporting** Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

**Evaluation Accountability Standards**

The evaluation accountability standards encourage adequate documentation of evaluations and a metaevaluative perspective focused on improvement and accountability for evaluation processes and products.

• **E1 Evaluation Documentation** Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

• **E2 Internal Metaevaluation** Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.
Appendix B Ethics Approval

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

February 17, 2010

Dr. Heather Stuart
Department of Community Health and Epidemiology
21 Arch Street, 3rd Floor
Queen's University

Dear Dr. Stuart,

Study Title: Developing Best Practices in Anti-Stigma Programming for High School Youth: A Secondary Analysis of Contact-Based Programs
Co-Investigators: Ms. Michelle Koller

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following list of ethics requirements you must fulfill over the course of your study:

➢ Reporting of Amendments: If there are any changes to your study (e.g., consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. (see http://www.queenu.ca/opr/ree.htm).

➢ Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

➢ Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

➢ Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair’s signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]
Chair, Research Ethics Board

Date

Study Code: EPID-303-10

➢ Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
QUEENS UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS
RESEARCH ETHICS BOARD (HSREB)

HSREB Renewal of Ethics Clearance

February 26, 2018

Dr. Heather Stuart
Department of Public Health Sciences
Queen's University

RESEARCHER NAME: JASON S. CLARK
Department Code: APH-003-19
Study Title: Developing Best Practices in Anti-Smuggling Programming for High School Youth: A Secondary Analysis of Content-Based Programs

Review Type: Delegated
Date Ethics Clearance Effective: March 26, 2018
Date Date of Initial Review: March 26, 2018

Dear Dr. Stuart,

The Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB) has reviewed the application. The study, including all currently approved documentation, has been granted ethical clearance until the expiry date noted above.

Prior to the expiration of your ethics clearance, you will be reminded to submit your renewal request through ROMEO. Any lapses in ethical clearance will be documented below.

Yours sincerely,

[Signature]

Chief, Health Sciences Research Ethics Board

The HSREB operates in accordance with the requirements of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), the International Conference of Harmonisation Good Clinical Practice Consolidated Guidelines (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the National Health Products Regulations; Part 1 of the Medical Devices Regulations; Canadian General Standards Board, current provisions of the Canadian Personal Health Information Protection Act (PHIPPA 2005) and its separate regulations. The HSREB operates through the COPP RED Qualifications Program and is registered with the U.S. Department of Health and Human Services (DHHS) Office for Human Research Protection (OHRP); Federale Assurance Number: FWA#00000514; HS#0000177.

HSREB members involved in the research project do not participate in the review, discussion, or decision.