THE ASSESSMENT OF THE QUALITY OF MENTAL HEALTH LITERACY MEASUREMENT TOOLS: A SCOPING REVIEW AND THREE SYSTEMATIC REVIEWS

by

Yifeng Wei

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

March 2017

© Copyright by Yifeng Wei, 2017

DEDICATION PAGE

This thesis work is dedicated to my husband, Bin Xie, who has been a constant source of support and encouragement during the challenges of graduate school and life. I also express my thanks and appreciation to Dr. Patrick McGrath who has generously given his time and expertise to guide my research work. I am grateful to my thesis committee members, Dr. Jill Hayden and Dr. Stan Kutcher, who shared their meticulous research and insights that supported and expanded my work.

Table of Contents

LIST OF TABLES vii

LIST OF FIGURES viii

ABSTRACT ix

LIST OF ABBREVIATIONS USED x

ACKNOWLEDGEMENTS xi

CHAPTER 1 INTRODUCTION 1

1.1 EPIDEMIOLOGY OF MENTAL ILLNESS 1

1.2 MENTAL HEALTH LITERACY 1

1.3 PREVIOUS RESEARCH ABOUT MENTAL HEALTH LITERACY 3

1.4 GOALS OF THIS THESIS 5

CHAPTER 2: MENTAL HEALTH LITERACY MEASURES EVALUATING KNOWLEDGE, ATTITUDES AND HELP-SEEKING: A SCOPING REVIEW [39] 7

2.1 BACKGROUND 7

2. 1. 1 Epidemiology of Mental Illness 7

2. 1. 2 Mental Health Literacy 7

2. 2 METHODS 9

2. 2. 1 Search Strategy 9

2. 2. 2 Inclusion Criteria 9

2. 2. 3 Exclusion Criteria 10

2. 2. 4 Data Extraction and Study Selection (Charting) 10

2. 3 RESULTS 12

2. 3. 1 Knowledge Measures 13

2. 3. 2 Stigma Measures 14

2. 3. 3 Help-Seeking Measures 17

2. 4 DISCUSSION AND FUTURE DIRECTIONS 18

2. 4. 1 Representative Samples 18

2. 4. 2 Geographic Weighting 19
2. 4. 3 Adequacy of Measurement 19
2. 4. 4 Knowledge Measures 20
2. 4. 5 Stigma Measures 21
2. 4. 6 Help-seeking Measures 22
2. 5 LIMITATIONS 22
2. 6 IMPLICATIONS AND CONCLUSIONS 23
2. 7 SUMMARY 24
CHAPTER 3: MEASUREMENT PROPERTIES OF TOOLS MEASURING MENTAL HEALTH KNOWLEDGE: A SYSTEMATIC REVIEW [40] 25
3.1 BACKGROUND 25
3. 2 METHODS 28
3. 2. 1 Search Strategy 28
3. 2. 2 Selection Criteria 29
3. 2. 3 Data Extraction 30
3. 2. 4 Study Quality Assessment (Risk of Bias Assessment) 30
3. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality 31
3. 3 RESULTS 32
3. 3. 1 Study Selection and Characteristics 32
3. 3. 3 Quality of Measurement Properties 36
3. 4 DISCUSSION 39
3. 5 LIMITATIONS 42
3. 6 CONCLUSIONS 43
CHAPTER 4: THE QUALITY OF MEASUREMENT TOOLS EVALUATIING THE STIGMA OF MENTAL IILLNESS: A SYSTEMATIC REVIEW [41] 44
4. 1 INTRODUCTION 44
4. 2 METHODS 45

4. 2. 1 Search Strategy 46
4. 2. 2 Selection Criteria 47
4. 2. 3 Data Extraction 48
4. 2. 4 Study Quality Assessment (Risk of Bias) 48
4. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality 49
4. 3 RESULTS 50
4. 3. 1 Study Selection and Characteristics 50
4. 3. 2 Methodological Study Quality 52
4. 3. 3 Level of Evidence on the Overall Quality of Measurement Properties 53
4. 4 DISCUSSION 55
4. 5 LIMITATIONS 59
4. 6 CONCLUSIONS 60
CHAPTER 5: MEASUREMENT PROPERTIES OF MENTAL HEALTH LITERACY TOOLS MEASURING HELP-SEEKING: A SYSTEMATIC REVIEW [42] 61
5. 1 INTRODUCTION 61
5. 2 METHODS 62
5. 2. 1 Search Strategy 62
5. 2. 2 Selection Criteria 63
5. 2. 3 Data Extraction 64
5. 2. 4 Study Quality Assessment (Risk of Bias) 64
5. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality 65
5. 3 RESULTS 66
5. 3. 1 Study Selection 66
5. 3. 2 Study Characteristics 67
5. 3. 3 Methodological Quality of Studies 68
5. 3. 4 Level of Evidence on the Overall Quality of Measurement Properties 69
5. 4 DISCUSSION 71

5. 5 LIMITATIONS 74

5. 6 CONCLUSIONS 75

CHAPTER 6: CONCLUSIONS 76

REFERENCES 81

APPENDICES 106

Appendix 1: Figures and Tables 106

Appendix 2: Search Strategies in PubMed 169

Appendix 3: Quality Criteria of Measurement Properties 172

Appendix 4: Level of Evidence for the Overall Quality of the Measurement Property 175

LIST OF TABLES

Chapter 2 Table 1 Psychometrics of knowledge measures 115

Chapter 2 Table 2 Psychometrics of stigma/attitudes measures 117

Chapter 2 Table 3 Psychometrics of help-seeking measures 124

Chapter 3 Table 1 Study characteristics 126

Chapter 3 Table 2 Methodological quality of a study on each measurement property of a measurement tool 131

Chapter 3 Table 3 Quality of each measurement property 133

Chapter 3 Table 4 Overall level of evidence of measurement properties 136

Chapter 4 Table 1 Study characteristics 139

Chapter 4 Table 2 Methodological quality of included studies, quality of each measurement property and overall level of evidence of measurement properties 147

Chapter 5 Table 1 Study characteristics 155

Chapter 5 Table 2 Methodological quality of a study on each measurement property of a measurement tool 162

Chapter 5 Table 3 Quality of each measurement property 164

Chapter 5 Table 4 Overall level of evidence of measurement properties 167

LIST OF FIGURES

Chapter 2 Figure 1 Charting process (data extraction process) 106

Chapter 2 Figure 2 Search results 107

Chapter 2 Figure 3 Study participants by study numbers 108

Chapter 2 Figure 4 Study site by study numbers 109

Chapter 2 Figure 5 Publication dates by study numbers 110

Chapter 2 Figure 6 Measure content by study numbers in each outcome 111

Chapter 3 Figure 1 Flow chart of search results 112

Chapter 4 Figure 1 Flow chart of search results 113

Chapter 5 Figure 1 Flow chart of search results 114

ABSTRACT

BACKGROUND

Mental health literacy interventions have received increasing attention as a strategy to promote positive mental health, improve early identification of mental disorders, reduce stigma and enhance help-seeking behaviors. However, despite the abundance of research on mental health literacy interventions, there is an absence of evaluations of current available mental health literacy measures. This research responds to this need through a scoping review and three systematic reviews on the scope and quality of mental health literacy measurement tools.

METHODS

We searched PubMed, PsycINFO, Embase, CINAHL, the Cochrane Library, and ERIC for relevant studies without limits on study participants, locations, or publication dates. Searches and analysis were conducted between 2013 and 2016. We included English publications of quantitative studies addressing psychometrics of mental health literacy tools. We hand-searched reference lists of included studies and additionally searched Google Scholar for additional studies. We assessed the methodological quality of included studies, the quality of each measurement property, and determined the overall level of evidence for measurement properties across studies.

RESULTS

We included 16 knowledge tools (17 studies), 101 stigma tools (117 studies), and 12 help-seeking related tools (24 studies) for assessment. We found that knowledge measures mainly investigated the ability of illness identification, and factual knowledge of mental disorders. Stigma measures addressed personal/perceived stigma against mental illness, self-stigma, experienced stigma; and stigma against mental health care. Help-seeking measures assessed help-seeking attitudes, intentions to seek help, and actual help-seeking behaviors. Thirteen mental health knowledge tools, 11 help-seeking tools, and 81 stigma measurement tools were rated as having “limited”, “moderate” or “strong” level of evidence. The level of evidence for the rest of the tools were considered as “conflicting” or “unknown”.

CONCLUSIONS

This research provides a compendium of available mental health literacy measures with their quality assessed. Future research may focus on the generalizability of the tools across diverse settings, follow standard guidelines to improve the quality of future psychometrics studies, and may develop and test mental health literacy interventions based on the findings of this research. However, the validation of measurement tools is an ongoing process and additional research may consolidate our recommendations.

LIST OF ABBREVIATIONS USED

ATHSS	Attitudes Towards Help-Seeking Scale
ATSPPH	Attitudes Toward Seeking Professional Psychological Help Scale
ATSPPH-SF	Attitudes Toward Seeking Professional Psychological Help Scale short form
Jorm MHL	Jorm Mental health literacy survey
IASMHS	The New Inventory of Attitudes Towards Seeking Mental Health Services
HSAS	Help-Seeking Attitude Scale
ASPH-S	Scale of Attitudes Toward Seeking Psychological Help for Secondary Students
HSA	Help Seeking Acceptability
PATPSI	Parental Attitudes Toward Psychological Services Inventory
MHLS	Mental Health Literacy Scale
ISCI	Intention of Seeking Counseling Inventory
GHSQ	General Help Seeking Questionnaire
HSI	Help Seeking Intentions
COSMIN	Consensus-based Standards for the Selection of Health Measurement Instruments

#

#

ACKNOWLEDGEMENTS

I would like to acknowledge that this study is supported by Yifeng Wei's Doctoral Research Award - Priority Announcement: Knowledge Translation / Bourse de recherché, issued by the Canadian Institutes of Health Research. Dr. McGrath was supported by a Canada Research Chair. In addition, I would like to express deep appreciation to Ms. Catherine Morgan, Michelle Xie, Su Zhang for their tremendous help with data collection and analysis, and the health librarian, Ms. Robin Parker, who helped with designing the search strategies of this research.

CHAPTER 1 INTRODUCTION

1.1 EPIDEMIOLOGY OF MENTAL ILLNESS

Mental disorders are common among young people between 12 and 25 years of age [1-5]. Worldwide, approximately 1/3 of the burden of illness in this group of young people can be attributed to mental disorders [4-7]. Untreated mental disorders may prevent people from achieving success in schools and at work; stop them from functioning well at home and in the community; and may lead to premature mortality and reduced life expectancy [8-11]. A recent international cross-sectional study in 17 countries demonstrated that untreated mental disorders also are associated with increased risk of a wide range of chronic physical conditions (e.g., heart disease, stroke, cancer, diabetes mellitus, hypertension, asthma, other chronic lung diseases, and peptic ulcer) [12].

However, despite the great burden of illness incurred by these conditions, research shows, worldwide, between 70%-80% of young people and adults do not receive the mental health care they need although effective treatments are available [13-16]. This may be due to numerous and varied reasons, including the lack of knowledge about mental disorders and their treatments, stigma against mental illness, limited access and other socio-economic or cultural barriers to mental health care [17-18]. A recent systematic review [19] of perceived barriers and facilitators for mental health help-seeking further indicated that perceived stigma and problems in symptom identification were among the most important barriers to mental health help-seeking.

1.2 MENTAL HEALTH LITERACY

Mental health literacy interventions may be appropriate approaches to address these barriers and therefore facilitate mental health help-seeking [20-22]. Mental health

literacy is both a derivative, and component of health literacy. Health literacy is an evolving concept from a highly focused construct based on functional literacy (reading and writing) to one that focuses on the need for individuals to develop the knowledge, attitudes and competencies they need to become effective agents in their own health care [23]. World Health Organization recognizes health literacy as "a stronger predictor of an individual's health status than income, employment status, education level and racial or ethnic group" [24]. And it has the potential to improve not only individual but also population health as it is designed to "decrease health inequities in populations, and enhance the operation of health systems and the development of health policy" [23] (page 154).

Like health literacy, the concept of mental health literacy is also evolving. It was first defined as "knowledge and beliefs about mental disorders which aid their recognition, management or prevention" [25] (p. 182). This definition focuses on the early identification of symptoms of mental illness and awareness of help-seeking resources. More recently, mental health literacy is conceptualized to include 4 domains: 1) understanding how to obtain and maintain good mental health; 2) understanding mental disorders and their treatments; 3) decreasing stigma against mental illness; and 4) enhancing help-seeking efficacy [23, 26-27]. And these four domains aim to address 3 inter-related outcomes: knowledge (knowledge of mental illness and positive mental health), stigma/attitudes, and help-seeking efficacy. This more recent definition of mental health literacy is consistent with the current construct of health literacy endorsed by the World Health Organization as an empowerment tool for people to participate in their health care [24]. According to Kutcher and colleagues [23]:

"This definition of MHL [mental health literacy] is an extension of previous constructs, is consistent with the evolving construct of HL (health literacy), includes the concept of stigma which has historically often been separately considered and extends Jorm's concept of self-help strategies to the wider construct of help-seeking efficacy. This evolving definition is based on considerable earlier refinements of our understanding of MHL and both a robust literature that well describes the inter-relationship between mental health knowledge and various types of stigma as well as recent stigma theory constructs where the lack of knowledge is considered to be a driver of prejudice (negative attitudes) that then influences behaviours (discrimination)." (page 155)

1.3 PREVIOUS RESEARCH ABOUT MENTAL HEALTH LITERACY

This thesis work builds on, is an extension of and contributes to the ongoing development of the important health domain of mental health literacy. Previous research can be categorized as falling into one of two domains: mental health literacy overall and technical components related to the study of mental health literacy.

Although not all research of this issue has come to a similar conclusion, overall the available evidence shows that improved knowledge about mental health and mental disorders, better awareness of how to seek help and treatment, and reduced stigma against mental illness at individual, community and institutional levels may promote early identification of mental disorders, improve mental health outcomes and increase the use of health services [20-22]. This evidence is further enhanced by six systematic reviews on the effectiveness of mental health literacy interventions [27-32] with three reviews [27, 28, 32] addressing multiple components (knowledge, attitudes/stigma, or help-seeking) of mental health literacy and three reviews [29-31] focusing on stigma reduction strategies.

We also found that a systematic review of reviews analyzed approximately 500 school mental health interventions, most of which addressed the promotion of positive mental health [33] and demonstrated similar but less robust findings primarily related to the impact of mental health promotion.

In addition to systematic reviews on the impact of overall mental health literacy discussed above, there were previously available four literature reviews specifically describing stigma and knowledge measurement tools [34-37]. For example, a review by Link and colleagues [34] discussed and categorized available stigma measurement tools within the context of labeling theory [38]. Another literature review [35] discussed one specific stigma measurement tool, the Internalized Stigma of Mental Illness scale, and its 47 versions and related reliability and validity in different contexts. A narrative review [36] identified 14 stigma measurement tools and categorized them according to different stigma theoretical models. Lastly, we also identified a narrative review of 13 mental health literacy measures, as defined within the mental health literacy framework by Jorm and colleagues [25, 37].

However, while numerous mental health literacy interventions have been applied and evaluated, and reviews on mental health literacy measurement tools have been published, there has been a lack of comprehensive understanding of currently available mental health literacy measures and there is limited research determining the quality of measurement tools that evaluate mental health literacy interventions. Thus, there exists a need to better understand characteristics, as well as strengths and weaknesses of existing mental health literacy measures and to help guide the appropriate application of evidence-

based measures and further shape future development of mental health literacy measures. This thesis responds to this need.

1.4 GOALS OF THIS THESIS

This thesis is divided into 6 chapters: Chapter 1 discusses the epidemiology of mental health and mental disorders, the concept of mental health literacy and its potential to improve individual health outcomes; and this chapter points out the research gap in assessing of the quality of current mental health literacy measurement tools. Chapter 2 is a scoping review that defines and describes available mental health literacy measurement tools and it sets the foundation for the research on the quality of mental health literacy measurement tools for the rest of the chapters. Chapter 3 reports a systematic review that assesses the methodological quality of studies on mental health literacy measurement tools that specifically address mental health knowledge and the quality of each measurement property of included tools, based on which it determines the level of evidence of overall quality of measurement properties of each tool; Chapter 4 is similar to Chapter 3, and it is a systematic review focusing on mental health literacy measurement tools addressing stigma of mental illness, and Chapter 5 is similar to Chapter 3 and 4, and it is a systematic review focusing on mental health literacy measurement tools addressing mental health help-seeking. Then in chapter 6, the last chapter of this thesis, we summarize findings from the first 5 chapters, make overall recommendations on the use of current mental health literacy tools based on current available evidence, and form future research priorities in this area.

This is a manuscript-based thesis, and therefore chapters 2, 3, 4, and 5 are each an independent manuscript submitted to different peer-reviewed journals for publication. Chapters 2 and 3 have been published in BMC Psychiatry [39, 40]. Chapter 4 has been

submitted to Epidemiology and Psychiatric Sciences [41]. Chapter 5 is "in press" with the Journal of Mental Health [42] and has been updated from the original manuscript submitted to the Journal of Mental Health to enhance the interpretation of the findings. The criteria to determine the level of evidence of overall quality is slightly different in chapters 4 and 5 (see details in chapters 4 and 5). This discrepancy is due to the evolving nature of the research methodology as chapter 5 was accepted for publication before chapter 4 was completed and submitted for publication. Further, because of the manuscript-based nature of this thesis, there are overlapping parts across chapters, such as the introduction section of chapters 2, 3, 4, and 5, the methodology section of chapters 3, 4 and 5. All references in this thesis were renumbered and modified from each individual original manuscript according to Dalhousie Thesis Format Guidelines to keep the format and reference numbers consistent across the thesis.

The author of this research (Yifeng Wei) defined the study scope; developed the study structure; collected and analyzed the data; drafted, revised, and finalized the manuscript for journal submission; and responded to the journal reviewers' comments for each manuscript –based chapter (chapters 2-5). The author holds the copyright for chapters 2, 3, 4 and 5.

CHAPTER 2: MENTAL HEALTH LITERACY MEASURES EVALUATING KNOWLEDGE, ATTITUDES AND HELP-SEEKING: A SCOPING REVIEW [39][1]

2.1 BACKGROUND

2. 1. 1 Epidemiology of Mental Illness

Approximately 70%-75% of adult mental health problems and mental disorders start to manifest during adolescence or early adulthood (12-25) [1-2]. Globally, mental disorders make up about 1/3 of the burden of illness in adolescence and young adulthood [4-8, 43]. Untreated mental disorders in adolescents and young adults are strong predictors of poor vocational achievements, problematic interpersonal and family functioning, as well as reduced life expectancy due to associated medical conditions, such as diabetes, heart diseases and stroke, respiratory conditions, and suicide [8-11]. However, despite the great burden of illness incurred by these conditions, research shows, worldwide, between 70%-80% of people do not receive the mental health care they need [13-15]. A recent systematic review [19] of perceived barriers and facilitators for mental health help-seeking indicated that perceived stigma and embarrassment, problems in symptom identification and a preference for self-reliance were the most important intra-personal barriers to mental health help-seeking.

2. 1. 2 Mental Health Literacy

Mental health literacy is a significant determinant of mental health and has the potential to improve both individual and population health [26, 28, 44]. Evidence shows that improved knowledge about mental health and mental disorders, better awareness of how to seek help and treatment, and reduced stigma against mental illness at individual, community and institutional levels may promote early identification of mental disorders, improve mental health outcomes and increase the use of health services [20-22].

[1] This chapter is published with BMC Psychiatry [39].

We conceptualize mental health literacy to include 4 domains: 1) understanding how to obtain and maintain good mental health; 2) understanding mental disorders and their treatments; 3) decreasing stigma against mental illness; and 4) enhancing help-seeking efficacy [26-27]. And, therefore, mental health literacy addresses 3 inter-related concepts: knowledge, attitudes and help-seeking efficacy. This definition is consistent with the current construct of health literacy defined and promoted by the World Health Organization as an empowerment tool helping people to more optimally participate in their own health care and improve their health outcomes [24].

We located a number of systematic reviews on the effectiveness of mental health literacy interventions [27-31]. In addition, we also located literature reviews of a more technical nature, describing stigma and knowledge measures [34-36]. Taken overall however, analysis of these reviews demonstrates that, there has been a lack of comprehensive understanding of current available mental health literacy measures. Such analysis is necessary to help evaluate the growing literature in the field of mental health literacy research. Thus, there exists a need to conduct a set of studies to help better understand strengths and weaknesses of existing measures and to help shape future development of measures. We conducted a scoping review, a systematic approach to map the literature in an area of interest and to accumulate and synthesize evidence available. This current scoping review was guided by Arksey and O'Malley's work [45], proposing four purposes: 1. to examine the extent, range and nature of research activity; 2. to determine the value of undertaking a full systematic review; 3. to summarize and disseminate research findings; and 4. to identify research gaps in the existing literature.

We analyzed available mental health literacy measures in general and those that focused on four common mental disorders with onset before or during adolescence and young adulthood: Schizophrenia, Depression, Anxiety Disorders, and Attention Deficit Hyperactivity Disorder (ADHD).

2. 2 METHODS

We used the definition of mental health literacy [26- 27] that is composed of 4 constructs addressing three outcomes: mental health knowledge (including knowledge about positive mental health (construct 1) and knowledge about mental illness and treatments (construct 2)), stigma/attitudes towards mental illness, and help-seeking, to define our search scope.

2. 2. 1 Search Strategy

One of the authors of this review and a health librarian designed the search strategies together. We searched PubMed, PsycINFO, Embase, CINAHL, Cochrane Library, and ERIC between 2013 and 2014, and re-ran the search in 2015. We applied four sets of search terms to identify domains of mental health literacy as outlined in Appendix 1.

2. 2. 2 Inclusion Criteria

We included quantitative studies that used, developed, or investigated measurement properties of mental health literacy measures evaluating any one, or combinations of the mental health literacy outcomes: knowledge, stigma/attitudes towards mental disorders, and help-seeking. Study designs included any type of quantitative studies: randomized controlled trials (RCTs), cluster RCTs, quasi-experimental studies; cohort studies; cross-sectional/survey studies, and controlled-before-and-after studies (pre/post tests). Only studies published in English were eligible

and non-English publications were excluded at the screening stage. Year of publication and study participants, including their age, were not restricted.

2. 2. 3 Exclusion Criteria

Studies were not eligible if they addressed mental health literacy but did not mention or describe the measure applied in the study. Studies of smoking prevention/cessation and other substance use prevention programs were not included. Studies of suicide prevention interventions that did not address related mental disorders, such as Depression were not eligible. Qualitative studies were excluded.

2. 2. 4 Data Extraction and Study Selection (Charting)

Two reviewers used the search strategy to independently search pre-identified databases. We first screened out irrelevant studies which mostly focused on stigma against HIV/AIDS, cognitive behavioral therapies, substance abuse/smoking, resilience scales, and clinical treatment related studies by reviewing titles and abstracts. We then imported the remaining studies, into RefWorks 2.0 database management software (2001) [46]. Duplicates were removed. We then screened titles and abstracts again and briefly scanned the full text to exclude studies not evaluating target outcomes. All studies that passed this exclusion process were included in the third stage of review for relevance by scanning title, abstracts and the full text for relevancy. At the next stage, we reviewed full-text articles for all the final included studies. Additionally, we added original studies that were referenced in included studies that cited their psychometric properties. We also checked the reference list of included studies for additional studies.

We applied “charting” techniques to conduct data extraction. In a scoping review, “charting” is a data extraction technique to synthesize and interpret data by “sifting, charting and sorting materials according to key issues and themes” (page 26) [45, 47].

The key themes we followed in the "charting" of our data are the three outcome measures: knowledge, attitudes, and help-seeking, which was the base of the data categorization. We also charted data by year of publication, study location, study type, outcome measures, and types of psychometrics examined (e.g., reliability, validity, and responsiveness/sensitivity to change). The detailed charting process for this review is depicted in Figure 1.

A data extraction form, developed in advance, was used for data extraction. We categorized studies into four types, based on the extent of how psychometric properties were investigated and reported in the study: validation studies with evaluating psychometrics (any type) and/or responsiveness/sensitivity to change as the major purpose of the study (coded as P); studies evaluating effectiveness of interventions or survey studies evaluating psychometrics (any type) and/or responsiveness/sensitivity to change of the outcome measures (coded as I/P or S/P); studies just reporting but not evaluating psychometrics and/or responsiveness/sensitivity to change of the applied tool (coded as I/? or S/?); and studies mentioning the measurement tool applied but not reporting psychometrics (coded as I or S), including studies that quoted psychometrics from other studies, but did not evaluate it in the current study. We then sorted and defined the data by measures on knowledge, attitudes/stigma towards mental illness, and help seeking respectively, listed authors who first applied the tool, and calculated the number of psychometrics studies for each outcome measurement. In addition, we collated all psychometrics studies in separate tables. Figure 1 illustrates this process.

Once this charting process was completed, we reviewed all included studies, developed and populated tables, and created charts and figures according to the above-

described typology in an Excel spreadsheet. To help ensure consistency in interpretation and validity of the final results, one of the reviewers read and charted all included studies (author of this thesis). Then the second reviewer checked all tables and compared and discussed the results with the first reviewer and they came to a consensus on the interpretation of the results. One methodology expert and two content experts were invited to help make the final decision when consensus was not reached between the two reviewers.

2. 3 RESULTS

Figure 2 presents the flow chart of the screening process and final included studies. A total of 401 studies were identified that met study criteria, including113 studies containing 69 knowledge measures, 307 studies containing 111 stigma measures, and 91 studies containing 35 help-seeking measures. Measures that modified and applied the concepts of the original ones were not counted as a new measure in our review. Out of the 401 studies, 130 validation studies reported and evaluated psychometrics (reliability, validity and/or the responsiveness/sensitivity to change) of the measures applied (P, I/P, or S/P), including14 knowledge studies (14 measures) (Table 1) [48-61], 102 stigma/attitudes studies (65 measures) (Table 2) [54-55, 58, 62-161], and 19 help-seeking studies (10 measures) (Table 3) [54, 120, 162-178]. These 3 tables summarized characteristics of validated studies, however we only listed authors who developed or first applied the measures although we included and summarized study results from other authors. Of these 130 studies, 5 studies also evaluated and reported responsiveness/sensitivity to change. Total number of studies for each sub-category may not necessarily match the total number of included studies because some studies tested more than one measurement tool in one study.

Study characteristics, such as study participants, locations, publication dates, and tool outcomes are reported in figures 3-6. Studies were conducted in 32 countries, with the United States of America as the most commonly studied site, followed by Australia and Canada. Study participants were mainly post-secondary students, especially students in psychology or related professions, followed by the general public, and mental health service users (e.g., patients and their families). Most of the studies (n=337) were published after the year 2000.

2. 3. 1 Knowledge Measures

The most widely used knowledge measures (by the number of studies in which the measure was applied) include the Mental Health Literacy Questionnaire (Jorm MHL) by Jorm and colleagues (1997) [25], Mental Health Knowledge Schedule (MAKS) [51], the World Psychiatric Association (WPA) "Open the Doors" (WPA-OD) questionnaire [179], Depression Literacy Scale (D-Literacy) [98], Knowledge about Schizophrenia Questionnaire (KASQ) [48], Schizophrenia Knowledge Questionnaire (SKQ)[180], and In Our Voices (IOV) knowledge measure [181].

The 69 knowledge measures evaluated general knowledge about mental health, knowledge on specific disorders such as depression, schizophrenia/psychosis, ADHD, and anxiety disorders (Figure 6). They used different approaches to measure knowledge. Some measures, such as those based on the approach by Jorm et al. (1997) [25] used the recognition of specific mental disorders (e.g., depression or anxiety) from the vignette description of symptoms. Other knowledge measures evaluated factual knowledge about mental illness with the true/false/don't know approach. This includes fact-based tests on terminology, prevalence, causes, diagnosis, etiology, prognosis, consequences, and course of illness; and knowledge about recognition, support, employment, treatment/help-

seeking/controllability, and recovery/coping, etc. [e.g. 48, 51-52, 60, 98, 179-183]. One tool addressed the ability to distinguish mental illness from neurological or somatic illnesses [e.g. 58]. There were a number of measures combining stigma knowledge and mental health knowledge [114, 184-187]. Finally, some were self-evaluation measures of extent of knowledge [e.g. 188-189].

Of the 69 measures, psychometric properties were reported for 26 (38%). And the rest of the 43 measures (62%) had no psychometric properties reported. Of 26 measures with reported psychometrics, 14 measures were evaluated for psychometric properties, including 2 measures for responsiveness/sensitivity to change [48, 50]. These 14 measures evaluated general mental health knowledge (6 measures), depression (4 measures), schizophrenia (2 measures), ADHD (1 measure), and anxiety disorders (1 measure) (Table 1). The rest of the twelve measures only reported but didn't evaluate psychometrics (internal consistency) and therefore we didn't include them in Table 1.

Most knowledge measures applied self-report multiple choice answers (true, false, I don't know/not sure), or vignettes with open-ended/closed questions [e.g. 190], or used Likert-scale statements as self-evaluation formats.

2. 3. 2 Stigma Measures

Of all the stigma measures, the most widely used measures (by the number of studies where the measure was applied) include the Social Distance scale (SD) [68]; Opinions about Mental Health Illness (OMI) [78]; Community Attitudes towards Mental Illness (CAMI, a modified version of OMI) [153]; Devaluation-Discrimination (DD) [124]; Depression Stigma scale (DSS/PPSM) [98]; Attribution Questionnaire (AQ) [80]; Internalized Stigma of Mental Illness (ISMI) [145]; and Perceived Dangerousness (PD) [125].

The 111 focus of the stigma/attitudes measures included: 1. stigma against mental illness or the mentally ill, such as social distance (the degree to which people are willing to accept the mentally ill in regular social life), personal stigma (participants' personal attitudes toward people with mental illness) and perceived stigma (participants' beliefs about others' attitudes about mental illness); 2. self-stigma; 3. experienced stigma by mental health service users; 4. stigma against mental health treatment, psychiatry, help-seeking, or mental health care facilities. Further, some measures evaluated stigma against specific mental illnesses, such as depression, anxiety, ADHD, and schizophrenia/psychosis. Eleven studies (7 measures) did not report what aspects of stigma were measured (Figure 6).

Social distance measures investigated issues such as a person's willingness to engage the mentally ill in the workplace and the community (e.g., employment, renting, being neighbors, marriage) [65, 88, 125, 127, 143, 191]. Similarly, measures evaluating stigmatizing experiences by the mentally ill focused on challenges people with mental illness experience in family and social life [71, 94, 111, 118, 145, 149].

Measures evaluating personal and perceived stigma covered areas such as authoritarianism, benevolence, mental hygiene ideology, social restrictions to the mentally ill, and etiology [78, 153]. Other measures evaluated components such as stigma related to illness prevalence, consequences, dangerousness/threat, treatment and recovery of mental illness, or the social/family life, social responsibilities, human rights, intelligence [55, 93, 98, 100, 124, 130, 135, 154]. In addition, there were personal and perceived stigma measures focusing on emotional/rejection responses, willingness to help, and disclosure concerns [63, 67, 76, 80, 105, 117, 143].

Self-stigma measures mostly evaluated cognition such as self-esteem, self-confidence, self-satisfaction/concurrence, self-blame; negative emotions such as low pride of oneself, shame, embarrassment, sense of inadequacy, inferiority to others, helpless, pressure; and behaviors such as withdrawal, fear of seeking help, and secrecy [66, 82, 102, 111, 132, 140, 155].

Measures examining stigma against treatment/help-seeking/mental health care/medical model/psychiatry addressed perspectives and emotions. For example, some measures evaluated stigma towards help seeking (e.g., help-seeking as personal weakness; people seeking help being less likeable, disturbed, posed risks to others, and should hide the fact of seeking help) [120, 156]. Other tools [74, 141, 151] investigated stigma toward psychiatry, for example, skepticism towards psychiatry; and stereotypes of psychiatrists, psychiatric hospitals, patients, and psychiatric treatments. Some tools measured emotional responses (e.g., fear, discomfort and embarrassment) to psychological services and mental health care [108, 163].

Eighty one (73%) articles on stigma tools reported on some psychometrics. Sixty five measures had evidence of reliability (e.g., Crobach's α; item-total correlations; KR-20; test-retest reliability; inter-rater reliability), validity (e.g, construct; concurrent; discriminant; convergent; predicative), or responsiveness/sensitivity to change (Table 2). Sixteen measures demonstrated only internal consistency, but none included discussions on how this was measured. Of these 81 measures, 48 evaluated stigma against the mental illness/ the mentally ill in general; 11 were self-stigma measures; 6 evaluated personally experienced stigma; and 12 evaluated stigma against mental health treatment

(psychological and pharmacological), psychiatry, help-seeking, or mental health care facilities. One tool did not specify what it measured.

2. 3. 3 Help-Seeking Measures

Of the 35help-seeking related measures, the most widely used are: Attitudes towards Help-Seeking Scale (later modified as Attitudes toward Seeking Professional Psychological Help Scale) (ATSPPH)[162, 166]; the mental health literacy questionnaire (Jorm MHL) that contains items on beliefs towards treatments [25]; General Help Seeking Questionnaire (GHSQ) [164]; and Intention of Seeking Counseling Inventory (ISCI) [192].

These help-seeking measures evaluating help-seeking intentions; beliefs or attitudes towards seeking psychological help for mental health problems or illness; beliefs towards mental health help or treatment in general; actual help-seeking behaviors; help-seeking efficacy (e.g. knowledge about where and how to find help, and who to find help from); self-reported ability to help others; or multiple components such as help-seeking intentions, help-seeking efficacy, and barriers for help-seeking (Figure 6).

Unlike measures of stigma against help-seeking described above, measures evaluating attitudes towards psychological help-seeking mostly addressed: recognition of need for psychological help; interpersonal openness; confidence in and trustworthiness of mental health practitioners [162]. Measures evaluating beliefs toward treatment mostly evaluated the perceived helpfulness, effectiveness or safety of various interventions [169, 193], or the myths of treatment [194]. One measure [195] added social norm items on perceived attitudes of others (e.g., friends, employer) on depression intervention.

Measures evaluating help-seeking intentions examined willingness, or preferences to seek help from different sources (e.g., friends, families, professionals, religion, or

spiritual healers [170, 190, 196-200]. One measure [190] further evaluated 3 extra dimensions of help-seeking intentions: talking to the listed sources; comfort level of talking to these resources; and helpfulness of these resources. Another tool measured intention levels for various emotional/behavioral challenges among college students [163]. Two measures didn't specify how intentions were measured [179, 201].

Measures addressing help-seeking behaviors evaluated whether help-seeking was sought, and if so, what type of help was sought (formal vs. informal) for both stressful events and mental illness [196, 199, 202, 203].

Ten measures had some psychometric evaluation such as internal consistency, reliability, factor analysis, construct validity, and criterion validity [162, 164, 166, 168-173, 175-178]. Details of the psychometrics of these 10 measures are presented in Table 3. The 10 measures with psychometrics addressed attitudes or beliefs towards help-seeking or treatments, and intentions for help-seeking (Table 3). Two measures reported the internal consistency of the tool [190, 199], but did not discuss how this were measured, and therefore were not included in the table. No psychometric properties were reported on measures of help-seeking behaviors.

2. 4 DISCUSSION AND FUTURE DIRECTIONS

We identified a number of significant issues for consideration. These are: 1) representativeness of study samples; 2) geographic weighting: 3) adequacy of measurement of mental health literacy (knowledge, stigma, and help-seeking).

2. 4. 1 Representative Samples

Almost half of the studies (n=185) were conducted among adolescents and young adults, particularly with post-secondary students (n=117) (Figure 3) mostly from health related professions, such as psychology, social work, and nursing. This raises cautions

about the generalizability of findings as participants are not representative of the general population.

Even within the context of postsecondary education, much less attention (only 9 studies) has been paid to the mental health literacy of educators, who are important role models and youth influencers in addressing mental health literacy [204]. Further research into mental health literacy should take these important factors into account.

2. 4. 2 Geographic Weighting

Research on the measurement of mental health literacy started as early as in late 1950'S but did not bloom until after 2000 (n=336; 84%) (Figure5). Most studies (Figure 4) took place in developed countries, especially the United States (n=170; 42%). Although there is ethnic diversity in the United States, the United States cannot be seen to represent other cultures. Moreover, different countries have different health systems and this may impact the implementation of mental health literacy approaches. For studies conducted in developing countries, authors either adapted existing measures, or used the conceptual framework from developed countries to create their measures, however, very few discussed the process of translation or the method of cultural adaptation. Therefore, the impact of important contextual factors, such as culture, ethnicity, geographic locations, education and health system, on mental health literacy and its measurement is currently unknown.

2. 4. 3 Adequacy of Measurement

Our analysis suggests that, out of three outcomes of mental health literacy (knowledge, attitudes and help-seeking), most measures evaluated stigma (n=111), followed by measures that evaluated knowledge (n=69), and a smaller number of help-seeking (n=33). Only a relatively small number of measures were validated in any way.

Secondly, widely used measures are often not validated. For example, the WPA mental health knowledge questionnaire was applied in 9 studies but no research has been identified to analyze its psychometric properties except for internal consistency.

Given the high proportions of un-validated measures being applied, it was difficult to determine the value of the study results and not possible to conduct cross-study comparisons of different interventions. There is a pressing need to validate these measures before their application.

With the measures that have been validated, there has been no research identified that appraised the quality of psychometric studies, and therefore, we were not able to recommend which measures are better than others. Further, given that the measures included in this review vary in their content, purposes and quality (measurement properties), more advanced research, such as systematic reviews is needed to locate evidence-based measures for use. Consensus-based Standards for the Selection of Health Measurement ,ı¹–ß°æ ˇı–¹ (COSMIN) [205] has been developed to serve this purpose and could be adapted for use in the comparative evaluation of mental health literacy measures.

Further, our review did not identify any measures addressing knowledge of positive/good mental health. Future measures should investigate knowledge on how to obtain and maintain good health as this now is recognized as an important component of mental health literacy.

2. 4. 4 Knowledge Measures

Our findings indicate that the diagnostic vignette approach is widely used as a measure of mental health knowledge. However, a recent study in which diagnostic vignettes were compared against non-diagnostic vignettes showed an inability of

participants to discriminate across “normal” and “ill” categories [206]. Further study to establish the validity of the diagnostic vignette evaluation approach as a measure of mental health knowledge is needed.

The myths and facts approach to measure knowledge has covered a wide range of aspects of mental health. However, we are unable to determine if there are different and developmentally appropriate knowledge components addressed at different points of the life-span among the current available measures.

2. 4. 5 Stigma Measures

The plethora of stigma measures, developed from numerous different ideological models (e.g., labeling theory [38]; attribution framework model [80]; cognitive behavioral model [207]; and social stigma model [208], has made evaluation of their validity in addressing stigma/attitudes challenging. The challenge has been to both validate each of the specific models and to determine which model may provide a better explanatory prediction for stigma or attitudes in different groups of people.

Further, only a few measures have targeted people’s emotional responses (n=8) towards mental illness. This is an important area because stigma is associated with self-experience of unpleasant feelings about mental illness and this may influence how people interact with those with mental illness [14]. Only very recently has research measured the stigma experience of people with mental illness (n=28 studies). This may provide a more comprehensive picture of how society treats people with mental illness. This may help to provide more concrete and useful information on how stigma interventions should be developed and delivered at both individual and community level.

Despite the challenges discussed above, this review has mapped out how stigma measures were developed and what they intended to measure, and this information may

provide researchers and practitioners some guidance on which path to take either in designing their measures, or applying/ adapting existing measures, or developing related interventions or programs in the future.

2. 4. 6 Help-seeking Measures

Help-seeking behaviors are challenging to measure as they are influenced by many factors, such as knowledge about the behaviors, attitudes and beliefs towards the behaviors, social norms, and intentions to perform the given behavior [209]. Most help-seeking measures in this review have focused on attitudes towards help-seeking/treatment (n=20) and intentions to seek help (n=11), and very few measures (n=4) directly measured actual help-seeking behaviors. Further, all 4 help-seeking behavior measures had no psychometric validation.

As Ajzen and Fishbein [209] pointed out, behaviors also may be influenced by self-expressed behavioral control which requires a person to have the skills, capacities, resources, and other important capacities needed to perform the behavior. However, we have not identified any measures to address these factors except for one tool measuring help-seeking efficacy (e.g. knowledge about where and how to find help, and who to find help from) [196].

2. 5 LIMITATIONS

We did not conduct a systematic review of the literature on available mental health literacy measures and therefore we are unable to come to conclusions about the quality of the studies applying the measures. We excluded non-English studies (n=21 at the title and abstract screening stage) and may have missed important measures in other languages. We did not check the grey literature that includes non peer-reviewed publication or documents/reports produced on all levels of governments and academics,

and therefore may have missed some eligible studies. We may also have mistakenly excluded some measures at the first screening stage of reviewing titles and abstracts where measures were not mentioned.

Additionally, although we tried to categorize and interpret measures within the category we attributed them to, some measures may contain items relevant to other categories, however we were unable to distinguish them with available information we have.

2. 6 IMPLICATIONS AND CONCLUSIONS

Our review provides a compendium of available mental health literacy measurement measures for researchers and practitioners who are interested in applying existing measures or developing new measures that of particular relevance to their work. Because of how we selected eligible studies, our review further automatically forms a comprehensive dataset of current mental health literacy interventions for stakeholders to consider for their use. This review also identifies the many gaps in the field, such as the unbalanced application of knowledge and help-seeking evaluation measures compared to the stigma/attitudes measures, the yet-to-be validated measures in each outcome category, and the lack of measures that measure all components of mental health literacy concurrently. This gap identification could potentially guide future research work in the field. Further, we have conducted a thorough summary and synthesis of the psychometrics properties of included measures, and clarified the need to further investigate the quality of the psychometrics studies. At this stage, most of the measures were created without consultation with the intended participants such as students, teachers, patients or health providers. Future work should focus on joint collaboration

across disciplines, between investigators and stakeholders and across more varied demographic and geographic groups. [2]

2. 7 SUMMARY

Chapter 2 summarized and categorized currently available mental health literacy tools and it set the foundation for us to further investigate the quality of available mental health literacy measurement tools in the following 3 chapters (3, 4, and 5). These chapters are systematic reviews to assess the quality of mental health literacy tools measuring mental health knowledge, stigma against mental illness, and mental health help-seeking (help-seeking behaviors, help-seeking intentions, attitudes towards help-seeking, and knowledge about help-seeking). These chapters critically analyzed the methodological quality of studies on psychometrics of available mental health literacy tools, assessed the quality of each measurement property of included tools, and further determined the level of evidence of the overall quality of their psychometrics across studies. Based on the findings of these systematic reviews, we then made recommendations for future research and the application of evidence-based tools.

[2] This is the end of the publication of Chapter 2.

CHAPTER 3: MEASUREMENT PROPERTIES OF TOOLS MEASURING MENTAL HEALTH KNOWLEDGE: A SYSTEMATIC REVIEW [40][3]

3.1 BACKGROUND

Mental disorders affect approximately 1 in 5 people [1, 3]. They are the leading cause of the global burden of diseases with the highest proportion of burden occurring in people aged 10-29 years [4]. Without appropriate treatment, they result in significant negative impacts on both short and long term social, economic and interpersonal outcomes as well as increasing risk for all causes of early age mortality, including suicide [9]. A recent international cross-sectional study in 17 countries further demonstrated that mental disorders are associated with increased risks of the onset of a wide range of chronic physical conditions (e.g., heart disease, stroke, cancer, diabetes mellitus, hypertension, asthma, other chronic lung diseases, and peptic ulcer) [12]. Effective treatments are available, but are uncommonly accessed by most youth with mental disorders [13-16, 210-211]. A recent systematic review found that barriers to receipt of mental health care include lack of knowledge about mental illness and stigma related to mental illness [19].

Mental health literacy has been considered as an effective approach to address these identified challenges and it is foundational for mental health promotion, early identification and treatment of mental disorders [28, 212-213]. Mental health literacy includes 4 components: 1) knowledge about how to obtain and maintain good mental health; 2) knowledge about mental disorders and their treatments; 3) decreasing stigma

[3] This chapter is published with BMC Psychiatry [40]

against those living with mental disorders; and 4) enhancing help-seeking efficacy [26]. Research shows that improved mental health literacy may be able to promote early identification of mental disorders, improve mental health outcomes, increase the use of health services, and enable the community to take actions to achieve better mental health [20-23].

Mental health literacy is a derivative of health literacy that evolved from functional literacy applied in health care environments addressing treatment adherence to a broader framework that further includes social and cognitive skills to improve and maintain good health and it is considered as an empowerment tool in social and political contexts [24]. According to the World Health Organization (WHO) [24], health literacy is a significant independent determinant of health: ‘‘a stronger predictor of an individual’s health status than income, employment status, education and racial or ethnic group.” (page 7).

Numerous mental health literacy programs have been developed over the last two decades. For example, a recent systematic review identified 27 studies evaluating the effectiveness of mental health literacy programs in the secondary school setting, in which 15 specifically addressed mental health knowledge about mental disorders, and the rest of studies focused on stigma and help-seeking behaviors [27]. Another systematic review of reviews analyzed approximately 500 school mental health interventions, most of which addressed the promotion of positive mental health [33]. Further, a meta-analysis of a particular mental health literacy intervention, Mental Health First Aid, has shown a positive impact on knowledge about mental disorders and help-seeking resources [32].

However, there is a paucity of evaluations of the tools to measure mental health literacy. For example, many mental health knowledge evaluation tools used in mental health literacy studies are varied in content, purpose, and quality, which may lead to non-comparable study results and increase risk of biased conclusions. Although sometimes the content of a mental health knowledge tool may be specifically designed to be somewhat different from another depending on the local community in which it is deployed, tools used must be of acceptable quality as the use of tools with poor quality may result in non-evidenced and unreliable results when evaluating the effectiveness of mental health literacy interventions or investigating mental health literacy levels in order to develop appropriate interventions in the community.

We conducted a scoping review to summarize and categorize currently available mental health literacy measurement tools, however, we did not synthesize information on the psychometric properties of the included tools or assess the quality of the evidence available [39]. In this chapter we conducted a systematic review to critically appraise the quality of studies evaluating the measurement properties of tools addressing knowledge about mental disorders, assess the quality of included measurement properties, and determine the level of evidence of overall quality of measurement properties of applied tools. Such a review will help researchers to identify what/how measurement properties of a mental health knowledge tool can be validated in a psychometric study. It will further help the research community to better choose appropriate tools to evaluate existing mental health literacy interventions or guide the development of new interventions.

3. 2 METHODS

We followed the protocol recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (http://www.prisma-statement.org/) to report findings. We adapted and applied the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist manual for the critical appraisal of studies [214] and we further applied quality criteria for embedded measurement properties developed by the same group of professionals [205]. COSMIN checklist is a robust tool developed specifically for systematic reviews on psychometric studies.

3. 2. 1 Search Strategy

We searched the following bibliographic databases: PubMed, PsycINFO, EMBASE, CINAHL, the Cochrane Library, and ERIC, using four sets of search terms from the scoping review [39], with the consultation of a health librarian between January and June 2015, and further updated and extended the search in Feb and March of 2016 to identify relevant studies. Appendix 1 is an example of the search strategies applied in PubMed. In addition, to ensure as much as possible that we would not miss relevant publications, we also searched Google Scholar, using the names of included knowledge tools identified from the search and finally, we also checked reference lists of included studies for additional studies. Two authors of this review are experts in mental health literacy field and they contributed to ensure that relevant studies were included.

Two people from the research team (one of them is the author of this thesis) applied an iterative process to independently screen titles (stage 1); titles of remaining studies to further exclude irrelevant studies, abstracts or brief scanning of full texts if

abstract reviewing was not sufficient to make decisions of inclusion (stage 2); and full texts of citations identified in the electronic literature search (stage 3). Reference check and Google Scholar search were conducted following these 3 stages of search. Following this, they met to compare their final included articles, and review and decide together the inclusion of articles one reviewer didn't include but the other reviewer did. A systematic review methodologist and two mental health professionals (also authors of this review) were available to guide the search and data analysis and help making final decisions on included studies.

3. 2. 2 Selection Criteria

We included any quantitative studies that evaluated measurement properties (reliability, validity or responsiveness) of mental health knowledge tools. Studies for inclusion had to report not only the psychometrics of the tool but also the statistical analysis used to evaluate the tool. We focused on tools that address mental health in general or common mental disorders that typically onset during adolescent years, including depression, anxiety, Attention Deficit Hyperactivity Disorder (ADHD) and schizophrenia. Our search did not restrict the publication dates or the age of participants.

We excluded studies addressing substance use disorder although it is common among youth, due to the fact that it covers a wide range of sub areas, and requires an independent research strategy beyond the scope of our current study. We excluded studies that were not published in English and those that only reported the psychometrics of tools but did not describe the statistical analysis used to evaluate the tools. For examples, many studies only reported the Chronbach's alpha but did not describe how this was achieved and therefore there were no data available for the quality assessment.

3. 2. 3 Data Extraction

We used the COSMIN checklist manual [205, 214] to develop a data extraction form. According to the COSMIN checklist [205, 214], a systematic review of studies on measurement properties could cover any of the following 9 areas in 3 dimensions. This includes: 1. Reliability (e.g. internal consistency, reliability (e.g. test-retest, intra-rater reliability, and measurement error); 2. Validity (content validity, structural validity (e.g. factor analysis), hypothesis testing (construct validity), cross-cultural validity, and criterion validity); and 3. Responsiveness (e.g. sensitivity to change). In addition, we followed the COSMIN checklist recommendation to document the population (e.g., age and gender), setting (e.g., country and culture), tool content and format, as well as types of psychometrics assessed in the included studies.

3. 2. 4 Study Quality Assessment (Risk of Bias Assessment)

We applied the COSMIN checklist with a 4-point scale [205, 214] to assess the methodological quality of each available study for each measurement property. The COSMIN checklist has 7-18 items to assess the study design and statistical methods for each property, with each item ranked as “excellent”, “good”, “fair”, or “poor” (see COSMIN checklist: http://www.cosmin.nl/). The overall methodological quality of each study assessing a measurement property is ranked as “excellent”, “good”, “fair, or “poor” by taking the lowest rating of any item in a box (worst score counts). For example, the domain for a study assessing the internal consistency contains 11 items for evaluation. If any one of the 11 items is scored “poor” but the rest of the 10 items are scored “excellent”, “good”, or “fair”, the final score for the study on internal consistency is “poor”.

3. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality

The level of evidence of the overall study quality of a measurement property was determined by the methodological quality of the available studies as determined by the COSMIN checklist stated above [205 214] and the consistency of the quality of measurement properties (positive (+), negative (-), indeterminate (?) findings) [216]. The details of the criteria for the quality of each measurement property can be found in Appendix 2. These criteria for the level of overall evidence were informed by Terwee and colleagues [205, 214] as refined in a systematic review of questionnaires measuring continuity of care [215] and Cochrane Back & Neck Group's recommendations on the overall quality of the evidence of each assessed outcome [216] (Appendix 3). As a result, the overall quality rating of a measurement property across studies was determined as one of the 5 levels of evidence: strong (+++ or ---), moderate (++ or --), limited (+ or -), conflicting (+/-) or unknown (x) (Appendix 3). The unknown (x) rating includes studies of poor methodological quality, as well as studies in which the quality of measurement properties was rated as "indeterminate" regardless of the study quality.

In March and April of 2016, two reviewers separately rated the quality of studies (one of them is the author of this thesis), the quality of each measurement property, and synthesized the levels of overall quality of measurement properties. Both reviewers studied and discussed the ranking system to make sure they were confident about its application. They compared and discussed their final rankings of the included studies and measurement properties. An Excel data ranking form was created for each level of analysis to store and keep track of quality scores for each reviewer. For rankings confirmation when they did not agree, a systematic review methodologist and two mental

health professionals (also authors of this review) were available to solve the differences between the two reviewers.

Based on the overall level of evidence, we considered measurement properties with strong positive ratings (+++) as ideal; moderate positive ratings (++) as preferred; and limited positive ratings (+) as minimally acceptable for use in research and practice. However, tools with measurement properties of negative ratings (---, --, -), or conflicting ratings (+/-), or unknown (x) have yet to be further studied before application since the quality of these properties was under the threshold or indeterminate defined by Terwee and colleagues regardless of the study quality [205, 214].

3. 3 RESULTS

3. 3. 1 Study Selection and Characteristics

Figure 1 demonstrates the flow chart of search results. The data were imported into Reference 2.0 database management software and duplicates were removed [46]. As described in methods section, we first checked study titles and screened out duplicates and studies unrelated to our topic of interest, such as studies measuring HIV/AIDS interventions, cognitive behavioural therapies, resilience programs, or knowledge about other specific mental disorders (e.g., post-partum depression, eating disorders, autism) which were not the focus of our current review. We further checked both titles and abstracts and screened out studies based on criteria in the first stage, as well as non-English publications. This procedure was repeated until the last stage of full text scanning and we excluded studies addressing other aspects of mental health literacy: stigma and help-seeking. As a result, we identified 131 studies that contained tools measuring mental health knowledge in which 17 studies provided psychometrics analysis of 16 tools

applied in these studies. Our analysis focused on the psychometrics of these 16 knowledge measurement tools, which are: Knowledge about Schizophrenia Questionnaire, Knowledge about Schizophrenia Test, Multiple-Choice Knowledge of Mental Illnesses Test, Mental Health Knowledge Schedule, Depression Multiple Choice Question, Depression Literacy, Anxiety Literacy, Test of Knowledge About ADHD, Knowledge about Depression and Mania Inventory, Journey of Hope Outcome Survey, Knowledge of Mental Disorders, Adolescent Depression Knowledge Questionnaire, Mental Health Disorder Recognition questionnaire, Mental Health Knowledge Questionnaire, Knowledge Questionnaire on Home Care of Schizophrenics, and Mental Health Literacy Scale [48-58, 60-61, 217-220]. This includes 2 studies [53-54] assessing Depression Literacy; another 2 studies assessing Knowledge about Schizophrenia Test [49, 218] and one study [54] evaluating 2 tools (Depression Literacy & Anxiety Literacy) in this current review.

We described the detailed study characteristics in Table 1. The 16 tools evaluated mental health knowledge among different populations: community members [49, 51, 60-61]; mental health patients [48, 52, 56]; patients' family members and caregivers [49, 56-57, 217-218]; police officers [49-50]; mental health professionals [49, 52, 218]; high school students [58, 60-61, 220]; post-secondary students [219]; athletes [54]; immigrants [53]; or elementary teachers [55]. The tools addressed either mental health knowledge in general [50-51, 57-58, 60-61, 219], or knowledge about specific mental disorders, such as depression [52-54, 56, 220], schizophrenia [48-49, 217-218], anxiety [54], and ADHD [55].

Fourteen tools focused on facts about mental illness, such as the etiology, diagnoses, prevalence, signs/symptoms, and comorbidity; as well as knowledge about treatments/side effects and mental health services [48-52, 54-57, 217-219]. Of these 14 tools, 1 (Mental Health Knowledge Schedule) further included stigma-related knowledge on help-seeking, recognition, support, and employment [51]; 1 (Knowledge about Depression and Mania Inventory) addressed knowledge about coping and illness management [56], and 1 (Knowledge about Schizophrenia Questionnaire) included knowledge about legal issues pertaining to mental illness [48]. Two tools (Knowledge of Mental Disorders, Mental health disorder recognition questionnaire) measured participants' ability to identify the illness appropriately [58, 60].

Table 1 indicates that 15 out of 17 included studies were conducted in Western countries with 35% of the studies conducted in the United States of (n=6), followed by Australia (n=3), United Kingdom (n=2), Canada (n=1), Germany (n=1), Italy (n=1), and Portugal (n=1). Two studies took place in non Western countries, China (n=1) and India (n=1). Study participants varied across studies and some studies included various types of participants, such as: family members of care givers of people with mental illness (n=5), community members (n=4), patients of mental illness (n=3), mental health professionals (n=3), police (n=2), high school students (n=2), university students (n=1), elementary school teachers (n=1), immigrants (n=1), and athletes (n=1).

3. 3. 2 Methodological Quality of Studies

Table 2 presents the methodological quality per study on each measurement property of a measurement tool. The 16 tools assessed properties such as internal

consistency (15 tools) [48-52, 54-58, 60-61, 218, 219-220], content validity (10 tools) [48-52, 55-56, 217, 219-220], construct validity (hypothesis testing) (7 tools) [49-50, 52, 56-58, 60], reliability (8 tools) [48-51, 54, 60, 219-220], structural validity/factor analysis (6 tools) [52, 55, 57-58, 61, 220], criterion validity (2 tools) [49, 56], responsiveness (sensitivity to change) (3 tools) [48, 50, 56] and cultural validity (1 tool) [218]. The methodological quality of included studies ranged mostly from "poor" to "good" (n=11) except that 5 studies addressing content validity [49-51, 56, 219], and 1 study [219] addressing internal consistency and structural validity demonstrated "excellent" quality. More than half (n=9) of the studies evaluating internal consistency were ranked as having "poor" quality while the rest were rated as "good" [52, 55, 57-58, 61, 220]. Studies evaluating reliability (n=8) also had mixed qualities ranging from "poor" to "good". Studies evaluating structural (n=6) and construct (hypothesis testing) (n=7) validity mostly demonstrated "fair" quality. All studies (n=3) examining responsiveness (sensitivity to change) were scored as having "poor" quality. One study was identified as assessing cultural validity with "fair" quality [218]. One study was identified assessing measurement errors with "good" quality [219].

Based on the quality criteria determined from use of the COSMIN checklist [214], study quality was downgraded if there were deficiencies of study design. For example, we found most (n=16) [48-58, 60-61, 217-218, 220] studies didn't report the percentage of missing items or described how missing items were handled, which may have introduced bias in their results [221], and therefore downgraded the study quality. Additionally, more than half of the studies (n=11) [48-51, 53-54, 56, 60-61, 217-218]

evaluated the internal consistency without checking unidimensionality of the tool resulting in "poor" quality of the study on this measurement property. The 2 studies [49, 56] evaluating criterion validity were rated as "fair" also due to the lack of justification regarding the "gold standard" the tool was compared against. Further, all studies evaluating construct validity (hypothesis testing) (n=10) [49-50, 52, 56-58, 60-61, 218-219] were rated as "fair" mostly because studies did not formulate the hypothesis "a priori", or the hypothesis was vague without specifying what was expected. And lastly, the "poor" quality of responsiveness (n=3) (sensitivity to change) [48, 50, 56] was mostly attributable to the application of inappropriate statistics such as effect sizes or *t-test* statistics.

3. 3. 3 Quality of Measurement Properties

While Table 2 presents the study quality, Table 3 presents the quality of each measurement property of all 16 tools. In terms of measurement properties by each tool (results by cases in the table), they all demonstrated mixed quality (+, -, or ?) as Table 3 demonstrated. When we investigated the quality by the measurement property (results by columns in the table), responsiveness received positive ratings (+) (above the quality criteria threshold) in all 3 studies it was evaluated [48, 50, 56]. The construct validity received positive ratings in all 8 studies it was evaluated [49-50, 52, 56-58, 60, 219], except that of 1 tool [60] with indeterminate (?) rating. The criterion validity evaluated in 2 studies [49, 56] demonstrated negative ratings (-) (below the quality criteria threshold). The rest of the measurement properties all demonstrated mixed ratings (+, -, or ?).

3. 3. 4 Levels of Evidence of Overall Quality of Measurement Properties

Table 4 demonstrates levels of evidence for the overall quality of each measurement property, which was determined by both the methodological quality of each study from Table 2 and the quality of each measurement property from Table 3. The criteria for the levels of evidence were developed to evaluate a measurement property of a tool in different studies. However, our review identified only 2 tools assessed in different studies [49, 53-54, 218], and the measurement properties for the rest of the 14 tools were assessed in only one study each. Therefore, the overall quality of these tools was based on 1 study only for each tool. Accordingly, two tools [60-61] demonstrated consistent positive ratings (+ or ++) (limited or moderate evidence) for their measurement properties. Two tools [48, 54] demonstrated unknown ("x") ratings for all measurement properties (studies of poor methodological quality or indeterminate quality of measurement properties). The rest of the tools showed mixed ratings (x, -, +, +/-, ++, --, +++, ---) of their measurement properties [49-58, 217-220].

In terms of overall ratings by measurement property (results by columns in the table), we found strong evidence (+++) of the content validity of 5 tools [49-51, 56, 218-219], and of the internal consistency of 1 tool [219]; moderate evidence (++ or --) of the internal consistency of 6 tools [52, 55, 57-58, 61, 220], of the content validity of 1 tool [217], and of the reliability of 2 tools [50, 219]; limited evidence (+ or -) of the reliability of 3 tools [49, 51, 60], the structural validity of 2 tools [52, 55], the criterion validity of 2 tools [49, 56, 218], and the construct validity of 9 tools [49-50, 52, 56-58, , 60-61, 219]. We also found the level of evidence of a number of measurement properties was unknown (x), including the responsiveness of 3 tools [48, 50, 56]; the internal

consistency of 8 tools [48-51, 54, 56, 217]; the reliability of 3 tools [48, 54]; the structural validity of 4 tools [57-58, 219-220]; the content validity of 4 tools [48, 52, 55, 220], and the measurement error of 1 tool [219].

According to the criteria in Appendix 3, the level of evidence of overall quality for a number of measurement properties was unknown "x" mainly because of poor study quality presented in Table 3, including the failure to assess the dimensionality of the tool which is the prerequisite for a clear interpretation of the internal consistency [222] and relatively small sample sizes (<30). Further, the level of evidence with negative ratings (- or --) was attributed to a number of factors, including the relatively weak correlations of two tools, the Knowledge about Schizophrenia Test and the Knowledge about Depression and Mania Inventory [49, 56] with gold standard tools (<0.70) when assessing the criterion validity; the lower-than-quality-threshold internal consistency (α<0.7) of Knowledge of Mental Disorders [58], or the failure of one study [55] on the tool Test of Knowledge About ADHD to discuss explained variance when assessing its structural validity.

Based on the level of evidence and criteria described above in the methods section, we recommend the application of 13 measures for their specific properties: Knowledge about Schizophrenia Test, Multiple-Choice Knowledge of Mental Illnesses Test, and Knowledge about Depression and Mania Inventory with their content (+++, Ideal) and construct (+, Acceptable) validity; Mental Health Literacy Scale with its internal consistency and content validity (+++, Ideal), reliability (++, Preferred), and construct validity (+, Acceptable); Mental Health Knowledge Schedule with its content

validity (+++, Ideal) and reliability (+, Acceptable); Depression Multiple Choice Question with its structural (+, Acceptable) and construct (+, Acceptable) validity; Test of Knowledge About ADHD with its internal consistency (+, Acceptable); Journey of Hope with its internal consistency (Preferred) and construct (+, Acceptable) validity; Knowledge of Mental Disorders with its construct (+, Acceptable) validity; Adolescent Depression Knowledge Questionnaire with its internal consistency (++, Preferred); Mental Health Disorder Recognition questionnaire with its reliability (+, Acceptable) and construct (+, Acceptable) validity; Mental Health Knowledge Questionnaire with its internal consistency (++, Preferred) and construct (+, Acceptable) validity; and Knowledge Questionnaire on Home Care of Schizophrenics for its content (++, Preferred) validity.

3. 4 DISCUSSION

This systematic review evaluated 16 mental health knowledge tools in 17 studies. It has provided a comprehensive critical analysis of the study characteristics, the methodological quality, the quality of individual measurement properties, and the overall evidence of the measurement properties of the included tools.

A review of the study characteristics indicates that most of the studies were conducted among the adult population and there were only four studies targeting youth [51, 53-54, 56]. This highlights the need for the development, evaluation and validation of tools addressing mental health knowledge specifically for youth who are at a vulnerable period of time related to the risk for developing mental illness. Further, most (n=15) studies were conducted in Western countries and cultural validity of the tools was

assessed in only one study. Therefore, at this time it is not possible to determine if measures created in one culture or setting can be appropriately used in another, especially in non-developed countries and regions where culture, social and economic contexts are dramatically different.

A strongly validated tool may not only help to accurately measure the impact of current mental health literacy interventions, but also can guide the development of new interventions. Rising from the assessment of study quality is the question of what constitutes a good psychometric study. Based on our findings and the COSMIN criteria, we propose that such a study may report on a sample size ≥30, examine the internal consistency and the dimensionality of the tool, determine the factors of the tool using factor analysis and explain the variances attributed to the factors, and establish the construct validity by testing pre-designed hypothesis. If it is a new tool, it is important to make sure tool items reflect the construct measured, are relevant to its population and fulfill its purposes. Also, such a study may examine the stability of the tool over appropriate period of time (usually 3 to 6 weeks). When a tool is applied in a culturally different setting, researchers may translate and back translate the tool, consider the adaption of the tool and pilot it in the target population (n≥10) before its application.

We recommended mental health knowledge tools by measurement properties because the level of evidence of each property within a tool was different even in the same study, and different tools measured different properties. Therefore, we decided it is not appropriate to conclude that one tool is better than the other. For example, the Mental Health Knowledge Questionnaire [61] was evaluated on two properties (internal

consistency and construct validity) and both reached the Acceptable and Preferred level of evidence. Another tool, the Mental Health Literacy Scale [219] was evaluated on six properties, four of which reached Acceptable or above level of evidence and two demonstrated level of evidence Unknown. In this case, we encourage readers to focus on the level of evidence of each individual property as well as their actual needs in practice when choosing which tool to use. Meanwhile, based on what we suggested above, researchers may further need to reach a consensus on what properties should be included for a psychometric study so that readers can compare the quality of different tools and make informed decisions.

However, as the validation of measurement properties is an ongoing and iterative process and needs to be conducted in different settings and contexts with different populations [223]. Further research could find that many of the measurement tools that demonstrated relatively low level of evidence of quality in the current review may have excellent psychometric properties with some populations in future research. More well-designed studies are needed to gather the evidence of the measurement properties to demonstrate their consistency and stability across studies.

The conceptual framework of mental health literacy includes 3 outcomes (knowledge, stigma and help-seeking), of which knowledge about positive mental health is a component. However, our review focused on tools addressing mental illness and we made this decision based on a number of factors. First, positive mental health covers a wide range of topics related to health promotion at individual, family, community and society level [224]. This includes social and emotional learning, resiliency, coping, social

and psychological welling, physical health, healthy eating, family relationship and connectedness, school and workplace environment, community involvement, and social support, to name a few. Each topic contains an independent and substantial body of research and unless we specifically come to a consensus on the scope and definition of each sub topic, it is unlikely that we are able to aggregate measurement tools in this area for use in assessments. Also, the mental health literacy concept is relatively new and the filter of each searched database is not sensitive to catch the search terms designed under the mental health literacy framework. We may have to design separate search strategies and conduct separate reviews to address this topic.

Lastly, as noted in the methods section, the COSMIN checklist applied the 'worse score counts' approach to determine the methodological quality of a property. This means a poorly scored item weighs more than all other well scored items in a criteria box. This may lead to a less positive score. For example, items in the criteria box for the content validity of DMCQ [52] were all rated as "excellent" on important factors such as constructs to be measured, purpose of the tool, and comprehensiveness of the tools, except one item rated as "poor" due to the failure to assess the relevancy of the tool for the study population. In this case, the final score of "poor" may not adequately reflect the true quality of the study.

3. 5 LIMITATIONS

We applied the COSMIN checklist originally developed to assess the quality of health status questionnaires and it may not be ideal for mental health knowledge tools in spite of some modifications that we made to the checklist. We didn't include studies

published in other languages, and therefore we may have missed some eligible studies. We only checked Google Scholar for grey literature because other available databases for grey literature such as GreyMatters (https://www.cadth.ca/resources/finding-evidence/grey-matters) is designed to contain information for health-related literature (e.g., health economics, clinical trials, drug and device information) and we decided they are not relevant to our topic of interest. However, this decision may have led to missing studies.

3. 6 CONCLUSIONS

To our knowledge, this review is the first to assess the quality of mental health knowledge measurement tools. We applied a standardized method, the COSMIN checklist, to evaluate quality of studies assessing measurement properties; we further assessed the quality of each measurement property, and provided a comprehensive and critical synthesis of current evidence in the field. The available evidence indicates that both the methodological qualities of included studies and the overall evidence of measurement properties are mixed. Based on the current evidence, we recommend that researchers consider using those knowledge assessment tools with measurement properties of positive ratings with strong and moderate evidence (++, or +++) or those with limited positive evidence (+) with caution (Table 4). However, our recommendation of specific tools was dependent on the context in which the tools were developed and validated. For example, the well-validated measurement property in one study may not be the same in another location or cultural context. Therefore, future research should focus both on improvements of current tools and their validation in different contexts.

CHAPTER 4: THE QUALITY OF MEASUREMENT TOOLS EVALUATIING THE STIGMA OF MENTAL IILLNESS: A SYSTEMATIC REVIEW [41][4]

4. 1 INTRODUCTION

Approximately 50% - 85% of people with severe mental disorders receive no treatment [6, 16]. Untreated mental illness results in numerous negative personal, economic, civic and social outcomes, including high rates of school dropout, poor vocational achievements, turbulent social interactions, and problems with the law [9, 225-226]. Untreated mental illness has become the dominant contributor to the global burden of non-communicable diseases [6-7].

People with mental illness have difficulty accessing effective mental health care due to a number of factors, amongst which stigma against mental illness is considered to be a significant barrier according to a recent systematic review on perceived barriers and facilitators to mental health help-seeking [19]. Stigma of mental illness was first defined as "a trait that is deeply discrediting that reduces the barer from a whole to a tainted, discounted one" [38]. Several conceptual frameworks have been created since, such as labeling theory [38], social attribution theory [80], cognitive behavioral modeling [207], and social stigma modeling [208], to both help understand and evaluate stigma related to mental illness, and guide stigma reduction interventions. As a result, the dimensions of the stigma of mental illness vary from one theory to another, and so do the stigma measurement tools created under different theories. More recently, the mental health literacy framework [23, 26] considers stigma reduction as one of its four core constructs and stresses the link between stigma reduction and the improvement of mental health

[4] This chapter is submitted to the Epidemiology and Psychiatric Sciences [41]

knowledge and the enhancement of help-seeking behaviors. Research, such as randomized controlled trials and longitudinal cohort studies [227-229] have demonstrated the effectiveness of interventions designed based on this approach.

Under these frameworks, a plethora of measurement tools have been developed to evaluate stigma of mental illness from different lenses. For example, a recent scoping review [39] identified 65 stigma measures and a narrative review [36] identified another 14, both of which categorized these tools according to different theoretical models. Another narrative review of stigma research on measurement tools collected more than 100 stigma measures informed by labeling theory specifically [34]. And one narrative review [35] discussed 47 different versions of a specific tool, Internalized Stigma of Mental Illness (ISMI), and summarized related reliability and validity of included versions in multiple countries. However, despite the abundance of stigma measurement tools, and stigma impact research using them, there has been little, if any, research identified to investigate the quality of currently available stigma measurement tools.

We conducted a systematic review to critically analyze the methodological quality of studies on psychometrics of available stigma tools and further to determine the level of evidence of the overall quality of their psychometrics across studies. Based on our analysis we then make recommendations for further stigma research and the application or ongoing development of these tools.

4. 2 METHODS

This review followed the protocol recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (http://www.prisma-statement.org/) [230] to report its findings. We conducted risk of bias analysis of included studies with the Consensus-based Standards for the Selection of Health

Measurement Instruments (COSMIN) checklist [214]; assessed the quality of each individual psychometric property, using criteria developed by the COSMIN group [205]; and then rated the level of evidence of overall quality included psychometric properties, considering both the study quality and the quality of psychometric properties. COSMIN checklist is a consensus-based checklist used to evaluate the methodological quality of studies on the measurement properties of health status instruments [231].

4. 2. 1 Search Strategy

We searched the following databases: PubMed, PsycINFO, EMBASE, CINAHL, the Cochrane Library, and ERIC databases for relevant studies between January and June 2015 and updated the search between April and May 2016, assisted by a local health librarian. To ensure our search covered all components currently considered to be related to stigma (as framed within the construct of mental health literacy), our search strategy covered all 3 outcomes of mental health literacy (knowledge, stigma and help-seeking) and we did not exclude studies that self-identified as focused on knowledge or help-seeking outcomes until the last stage of data extraction because some mental health literacy measures include all three components. We applied the search strategy from the scoping review [39] that contained 4 sets of key words and phrases with regards to general mental health and mental disorders, 3 outcomes of mental health literacy, assessment tools, and study designs. Appendix 1 provides details of all search words and phrases applied searching PubMed.

Two team members independently searched the citations identified from database searches in 2015 (between January and June) and updated the search in April and May of 2016 for relevant studies. Both members followed the same procedures to assess potential relevance of studies: reviewing titles in general (stage 1), reviewing titles and scanning

abstracts (stage 2), briefly scanning full papers (stage 3), and reading full papers for data extraction (stage 4). Following these stages, we checked the reference list (citation search) of each included study for additional studies and further searched narrative reviews on stigma measurement tools for additional studies [34-36]. The two reviewers discussed their identified studies at the end of search phase and reached consensus on the final inclusion of studies. Other research team members who are experts in mental health and/or research methodology were available to solve any discrepancies on the final decisions for included studies.

4. 2. 2 Selection Criteria

We included any type of studies that assessed and reported any psychometrics (reliability, validity, and responsiveness) of a stigma measurement tool. Based on our understanding of conceptual frameworks on stigma of mental illness discussed above, we defined a stigma measurement tool as one that evaluated: personal or perceived stigma, experienced stigma, emotional responses to mental illness, and self-stigma of mental illness. Our search focused on tools addressing stigma of mental illness in general or stigma against common specific mental illnesses: Anxiety Disorder; Depression; Attention Deficit Hyperactivity Disorder (ADHD); Schizophrenia. For a study to be included in the review, it had to report not only the psychometrics of the tool, but also the statistical analysis of these psychometrics. We searched databases for studies published in English and did not limit the date of publication, or study participant age in our search.

We excluded qualitative studies and studies that only provided psychometrics of the tool applied but did not report the statistical analysis of these psychometrics. For example, many studies evaluating anti-stigma interventions reported the internal

consistency of the tool applied but didn't describe the statistical analysis related to it and therefore were excluded from our review.

4. 2. 3 Data Extraction

We followed the COSMIN checklist manual [214] and created a data extraction form a priori to document basic information of each included study, such as author information, the tool content, the response option of the tool, population, location of the study, and study sample size. We further documented information about measurement properties and categorized them according to the COSMIN checklist into: 1. reliability (internal consistency, reliability (test-retest and intra-rater reliability), and measurement errors; 2. validity (content validity, structural validity (factor analysis), hypothesis testing (construct validity), cross-cultural validity, and criterion validity); and 3. responsiveness (sensitivity to change).

We categorized tools that adapted other existing tools by adding/reducing items or changing original items as separate tools. However, if a tool was created in one study but in another was assessed for its factors and the number of final items was adjusted from the original tool due to the factor analysis, we considered them as the same tool as this is part of the usual ongoing process of finalizing scales. Therefore, we expected more stigma measurement tools to be found in this chapter than in Chapter 2 because the scoping review categorized tools by stigma concept only, regardless of whether these tools had the same items or not.

4. 2. 4 Study Quality Assessment (Risk of Bias)

We determined the quality of a study for a particular measurement property and rated each as: "excellent", "good", "fair", or "poor". As a study may assess more than one measurement property, a study may have multiple levels of quality for different

measurement properties it assesses. Our quality criteria followed the recommendations by the COSMIN checklist manual [214, 231], according to which there are 7-18 criteria items to assess the methodological study quality for each measurement property, rated as "excellent", "good", "fair", or "poor" under each item respectively. The final ranking of the study quality for each property takes the lowest criteria ranking (worst score account). For example, for the study on the structural validity (factor analysis), the COSMIN checklist contains 7 criteria items to assess the study quality, and if under each item the study has different ranking ranging from "poor" to "good", the final ranking for this study would be "poor" for structural validity.

4. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality

In addition to the study quality of each measurement property, the COSMIN group further developed quality criteria for each psychometric property (except for cross-cultural validity) [205]. Each property must reach a quality threshold to receive a positive rating (+), otherwise a negative rating (-), indeterminate rating due to the lack of data (?), or conflicting rating (+/-) if the findings are contradictory. Appendix 2 provides details of the criteria and related rating scales. Based on both the methodological study quality described in the above section and the quality of the psychometric property, we determined the overall quality of a psychometric property when this property is assessed in one study or multiple studies. The ratings were determined by adapting and applying criteria from a systematic review on measures of continuity of care [215] and the Cochrane Back and Neck Group's recommendations on the overall level of evidence of each assessed outcome [216] (Appendix 3). As a result, the levels of evidence are: strong (S) (+++ or ---), moderate (M) (++ or --), limited (L) (+ or -), conflicting (C) (+/-), or unknown (U) (x). We considered measurement properties with positive strong evidence

(+++) as “ideal”, moderate positive evidence (++) as “preferred”, and limited positive evidence (+) as “minimum acceptable”.

If a property is assessed in two studies and study quality reached “fair” or above and the quality of the measurement property is positive (+) in both studies, we used the “worst score” approach for the level of evidence, otherwise we determined the level of evidence as conflicting (C(+/-)). If a property is assessed in more than two studies and we found fair, good or excellent study quality in more than half of the studies (>50%), we considered the level of evidence as strong, moderate or limited, using the “worst score account” approach. For example, if a measurement property is rated as (+) or (–) consistently in studies with the mixed study quality of excellent, good and fair, the final rating is limited level of evidence (L(+) or L(-)). For the rest of the cases, the level of evidence is conflicting (C (+/-)).

We defined the level of evidence as unknown (U(x)) in the following cases: 1. if a property is assessed in one study only and the study quality is “poor”, or the psychometric property is indeterminate (?), 2. if a property is assessed in two studies, and the study quality is poor or property is indeterminate (?) in both studies, 3. If a property is assessed in more than two studies, and the study quality is poor or property is indeterminate (?) in half or more than half of the studies.

4. 3 RESULTS

4. 3. 1 Study Selection and Characteristics

Figure 1 presents the flow chart of study selection process. The data were imported into Reference 2.0 database management software [46] and duplicates were removed. We then screened 21089 studies, and excluded studies that were not the topic of interest at this stage (e.g., studies addressing HIV/AIDS stigma, Cognitive Behavioral

Therapy, resilience, social and emotional learning, mental disorders that were not the topic of interest of this review). We then used the same criteria to further screened out irrelevant studies until the last stage of screening when we excluded studies measuring mental health knowledge or mental health help-seeking as well as studies containing stigma measurement tools without statistical analysis of related psychometrics. As a result, we identified 117 studies that reported and analyzed psychometric properties of 101 stigma measurement tools [54-55, 58, 62-67, 69-113, 115-122, 124-127, 129-130, 132-161, 232-250]. Table 1 provides the details of the study characteristics of included studies. We classified tools into 5 categories according to what they measured: perceived/personal stigma against mental illness or the mentally ill; perceived/personal stigma against mental health care (e.g., treatment, help-seeking, mental health institutions or psychiatry as a profession); people's emotional responses to mental illness; experienced stigma by people with mental illness or their relatives/caregivers; and self-stigma by people with mental illness. We didn't categorize tools under a specific stigma theory because most were developed with combined components from various theories or based on interviews with target population.

Ninety-one out of 101 tools applied Likert-scale response format asking participants to rate the level of agreement on items addressing stigma (Table 1). The other 10 tools applied formats such as multiple choices (e.g., yes/no/don't know) (Personal Rejection Scale, Social Distance Revised, Knowledge Test of Mental Health, Client Attitude Questionnaire) [108, 135, 139-140]; responses on a100 mm visual analogue scale (Depression Attitude Questionnaire and General Attitude Questionnaire) [69, 237, 121]; error-choice response (Test of Knowledge about ADHD) [55]; open-

ended questions (Labeling scale) [63]; and prevalence and frequency of stigma experience (Stigmatizing Experiences Scale) [149].

Study participants were mostly people with mental illness (n=36) and their relatives and caregivers (n=6), followed by community members/general public (n=20), health care providers and staff (n=20), college students (n=15), secondary school students (n=8), and people from other professions such as educators (n=2), police (n=1), athletes (n=1), employers (n=1), and military personnel and veterans (n=1). Some studies used multiple groups of participants mentioned above (n=8).

Most studies took place in developed countries with the United States of America as the most studied site (n=44), followed by the United Kingdom (n=21), Canada (n=8), China (n=8), Australia (n=6), German (n=3), Sweden (n=3), Italy (n=3), Greece (n=3), Belgium (n=2), and Austria (n=2). The rest of the studies were conducted in one of the following countries: Norway, Netherlands, Finland, Japan, Ireland, Jamaica, Poland, Israel, India, South Africa, Ethiopia, and Jordan. Four studies were conducted in multiple countries.

4. 3. 2 Methodological Study Quality

Table 2 summarizes all the study quality data with 4 rankings: “excellent” (E), “good” (G), “fair” (F), or “poor” (p). Each study demonstrated mixed quality from “poor” to “good”, when addressing different measurement properties of a tool, except one study on the Generalized Anxiety Stigma Scale (GASS) demonstrating “good” or “excellent” study quality for all measurement properties assessed [100].

Of 117 studies, a total of 5 met criteria for “excellent” quality, measuring the internal consistency of Stigma-Devaluation scale [83], construct (hypothesis testing) and structural validity of Generalized Anxiety Stigma Scale [100], as well as content validity

of Opening Minds Scale for Health Care Providers, Self-Stigma Scale, and revised Discrimination and Stigma Scale [116, 242, 249].

"Good" quality studies were mostly those measuring internal consistency (n=67) (Table 2), followed by 5 studies on the content validity [88, 101, 113, 141, 149], 1 study on test-retest reliability [100], 1 study on hypothesis testing (construct validity) [246], and 1 study on structural validity [158].

Studies of "fair" quality were found in most studies evaluating structural validity (89 out of 93), construct validity (hypothesis testing) (85 out of 92), test-retest reliability (38 out of 45); as well as in most studies evaluating cross-cultural validity (3 out of 4) [107, 126, 152], and all studies (n=7) evaluating criterion validity [94, 106, 112, 115, 135, 155-156]. We further identified studies of "fair" quality in some studies evaluating internal consistency (n=5) [30, 40, 72, 120, 136], and content validity (n=8) [35, 83, 85, 92, 95, 102,129, 241].

No studies on structural validity and criterion validity were identified as of "poor" quality, however the only two studies [113, 137] on the responsiveness of related tools were rated as "poor". We also found 36 studies with "poor" quality in evaluating the internal consistency of related tools. "Poor" study quality was further found in some studies evaluating content validity (n=10) [74, 84, 93, 112, 134, 153, 155-157, 161], test-retest reliability (n=5) [76, 54, 139, 145, 160], construct validity (hypothesis testing) (n=5) [77, 96-97, 112, 130], and cross-cultural validity (n=1) [83].

4. 3. 3 Level of Evidence on the Overall Quality of Measurement Properties

As described in previous sections, the study quality (E, G, F, or P) and the quality of measurement property (+, -, +/-, or ?) were combined to determine the level of evidence as: strong (S) (+++ or --), moderate (M) (++ or --), limited (L) (+ or -),

conflicting (C) (+/-), or unknown (U) (x), as shown in table 2. The quality of each measurement property helped to determine the direction of the level of evidence of overall quality as positive or negative and their ratings were presented in table 2 as well.

Our analysis found that the following measurement properties were considered to have strong evidence (+++) among 4 tools: the content validity of the revised Discrimination and Stigma Scale [249], Opening Minds Scale for Health Care Providers [116], and Self-Stigma Scale [242]; the internal consistency, structural validity (factor analysis) and construct validity (hypothesis testing) of the Generalized Anxiety Stigma Scale [100].

Moderate level of evidence (M(++); M(--)) were mostly the internal consistency of related tools (55 tools in 63 studies) (Table 2), as well as the content validity of 4 tools: revised Depression Attitude Questionnaire, Reported and Intended Behaviour Scale, Libertarian Mental Health Ideology scale, Stigmatizing experiences scale [88, 101, 141, 149]; test-retest reliability of Generalized Anxiety Stigma Scale [100]; and structural validity of Adolescent Attitudes toward Serious Mental Illness [158].

We further found limited level of evidence (L(+); L(-)) for construct validity (hypothesis testing) of 55 tools in 68 studies, structural validity of 46 tools in 56 studies, test-retest reliability of 23 tools in 29 studies (Table 2). This level of evidence was also found in the content validity of 8 tools [70, 83, 85, 92, 95, 102, 113, 129, 241], criterion validity of 7 tools [94, 106, 112, 115, 135, 155-156], and internal consistency of 2 tools [103, 233].

We identified conflicting (C(+/-)) evidence for the test-retest reliability of 9 tools [71, 74, 82, 117-118, 121, 130, 135, 145], the internal consistency of 6 tools [72, 80, 96-

97, 104, 108, 110, 126, 147, 153, 245], the construct validity of 5 tools [72-73, 79-80, 92, 108, 113, 124-126, 143, 159, 242], and the structural validity of 3 tools [72, 79, 126, 138, 153, 159, 242, 244, 250].

In addition, we were unable to determine the level of evidence for a number of measurement properties (U(x)) of some tools due to the lack of information provided. This includes the internal consistency of 29 tools in 37 studies, the structural validity of 25 tools in 26 studies (Table 2). We also found level of evidence unknown (x) for the content validity of 11 tools [74, 84, 93-94, 112, 134, 153, 155-157, 161], construct validity of 9 tools [63, 67, 73, 77-78, 90, 112, 117, 130], test-retest reliability of 4 tools [54, 76, 139], and responsiveness of 2 tools [113, 137]. There are 4 studies [83, 107, 126, 152] analyzing the cross-cultural validity of 4 tools, however, the COSMIN checklist has not developed criteria for the quality of this property, and therefore their level of evidence was scored unknown as well (U(x)).

Of 101 tools, 12 met the criteria of limited (minimum acceptable), moderate (preferred), or strong (ideal) positive level of evidence on all their assessed measurement properties (highlighted with ** in Table 2) [85, 105, 120, 127, 129, 146, 148, 151, 249, 158, 232, 234], and there are 69 tools met these criteria for some of their assessed measurement properties (Table 2). The rest of the 20 tools (highlighted with ?? in Table 2) did not reach at least the minimum acceptable level of evidence for related measurement properties..

4. 4 DISCUSSION

This review is the first of its kind to investigate the quality of studies containing tools evaluating stigma related to mental illness, as well as the quality of measurement properties of each included tool. As discussed above, a total of 81 tools met the criteria of

minimum acceptable, preferred, or ideal level of evidence with positive ratings for all or some of their measurement properties. These results may be useful for researchers and community members to consider for application in practice. Tables 1 and 2 can serve as a comprehensive resource for them to extract information they need, whether it is to design stigma reduction interventions, to develop new stigma measures, or to make decisions related to stigma of mental illness in the community.

However, it is a challenge to conclude one tool is better than the other for a number of reasons: 1. included tools contained different items addressing various domains of stigma, even for tools developed under the same theoretical framework; 2. studies evaluated different measurement properties; and 3. study quality and level of evidence varied even in the same study depending on the properties measured. For example, Attitudes to Severe Mental Illness measured general attitudes of the general public and is one of the 12 tools of which all measurement properties reached "limited" or "moderate" level of evidence [129]. Another tool, Reported and Intended Behaviour Scale [88] also measured general attitudes of the general public in multiple studies and had mixed level of evidence from "unknown" (x) to "moderate" (++). In this circumstance when choosing which tool for application, evidence of each individual property matters and we should also consider whether the purpose of the chosen tool (e.g., the content of the tool, target population, and the setting) is consistent with our actual application, either in developing an anti-stigma intervention or to measure public stigma of mental illness.

Yet, we don't recommend tools with negative ratings (---, --, or -) because the statistics of these measurement properties were below the criteria threshold, nor are we

confident about the application of tools with conflicting (+/-) or unknown (x) evidence. We also however raise the caveat that future recommendations on the use of these tools may change as we know that the validation of a tool is an ongoing process [223] and as more studies are conducted with more appropriate designs, tools that currently do not meet our criteria may do so following further future research.

In addition, the finding that there are currently over 100 different stigma measurement tools applied across 117 different studies raises concerns about the overall value of this body of research, as it is simply not possible to come to general considerations about issues related to stigma in mental illness given the use of so many different tools to measure the concept. As such, we were unable to decide which tool is the "gold standard" in this area and this is probably why only 2 [94, 156] out of 7 studies measuring criterion validity showed significant correlations with the pre-defined "gold standard" tools. Future research should focus on using a much smaller number of tools, those with the best psychometric properties to help decrease the uncertainty arising from the application of so many different tools of varying quality.

The study characteristics of these included validated tools are consistent with findings from the scoping review [39] that there are few tools (only 6 validated tool) assessing people's emotional responses to mental illness. Further, most research using stigma measurement tools was conducted in the United States of America and it is not known if tools applied this population can be compared to those applied in other countries. Similarly, there are few tools validated among secondary school students (n=8) and teachers (n=2), indicating a substantial contrast against the fact that most mental

disorders onset between the age of 12 and 25 [210] and most young people attend school during this period of time.

Measuring stigma against mental illness is challenging because of social desirability bias where people tend to answer questions in a manner that will be viewed favorably by others [251]. This bias may seriously jeopardize the validity of findings when the tool is applied. We found that only1 out of the 101tools addressed this potential bias by applying error-choice response for the Test of Knowledge about ADHD [73]. Future application of stigma tools may need to address this and consider evidence-based approaches to reduce social desirability bias. Some recommended techniques include the integration of social desirability scale assessment into the stigma assessment tool, the application of random response techniques, the addition of disguising of scale intent, or an indirect questioning approach [223].

Based on our findings and informed by the COSMIN checklist, we have recommendations for researchers to consider when designing psychometric studies. First, psychometric studies need to obtain an adequate sample size, and address missing items for relevant measurement properties. In addition, checking unidimensionality of items is as important as reporting Cronbach's alpha or KR-20 in deciding the study quality of internal consistency. Further, in examining test-retest reliability, the analysis on the independence of the test administration, the appropriate timing between tests, and the stability of test conditions were often ignored but matter in improving study quality. When assessing content validity, piloting the items in the targeting population ($\geq$10) for comprehensiveness is equally important as item selection process. In analyzing the structural validity/ factor analysis, it is essential that researchers report the variances

explained by factor analysis to improve study quality. When measuring construct validity, it is suggested that studies formulate hypotheses in advance and pre-define the direction and the magnitude of the mean difference or correlations of related statistical analysis to ensure the appropriateness of analysis.

It is also noted that the most assessed measurement properties were internal consistency, structural and construct validity in this review while responsiveness was the least studied property and measurement errors were not assessed by any included studies. Rising from this analysis is the question of what and how many psychometric properties should be included for psychometric analysis. Although the COSMIN checklist established criteria for 9 properties, it is a modular framework that doesn't require the evaluator to complete analysis of all 9 properties. However, informed by the findings from this current review, it is reasonable to propose that the validation of a tool should at least analyze whether: the tool items are appropriately related (internal consistency); it is reliable over time (test-retest reliability); the tool constructs are adequately established (structural and construct validity). Further when it is applied in culturally different settings, cross-cultural validity has to be evaluated prior to its application. The lack of cross-culturally validated tools makes cross cultural conclusions about stigma related to mental illness difficult if not impossible. This need requires further investigations by researchers in this field.

4. 5 LIMITATIONS

Our review is limited in excluding non-English publications (25 non-English potentially relevant citations were identified at the title and abstract screening stages) and therefore may have missed some eligible studies otherwise. Secondly, the COSMIN checklist may not be the most appropriate critical appraisal approach although it is the

only available one as it is originally designed for health status questionnaire. We have revised it for the analysis of included studies in this review, however may still have created unknown bias against stigma measurement tools included in this review. Further, the absence of grey literature may have also resulted in some missing studies although available databases such as GreyMatters [252] may not be so relevant as it mostly contains health-status information such as health economics, clinical trials, drug and device information.

4. 6 CONCLUSIONS

This is the first systematic review to investigate the study quality and overall level of evidence of tools evaluating stigma of mental illness. Our findings provide rich evidence on the psychometric properties of current stigma measurement tools so that researchers and decision makers can choose best available tools for use in practice. However, no matter what tools researchers or decision makers choose, it is recommended that researchers continue to validate tools in different settings to ensure that these tools are able to be appropriately used in numerous different contexts and populations.

CHAPTER 5: MEASUREMENT PROPERTIES OF MENTAL HEALTH LITERACY TOOLS MEASURING HELP-SEEKING: A SYSTEMATIC REVIEW [42][5]

5. 1 INTRODUCTION

Health literacy, including mental health literacy, has been considered as a significant predictor of positive health outcomes by the World Health Organization [24]. Mental health literacy focuses on knowledge and strategies to obtain and maintain good mental health, knowledge about mental disorders and related treatments, strategies to decrease stigma, and enhancement of help-seeking efficacy [26]. This framework highlights the importance of help-seeking related outcomes such as help-seeking behaviours, intentions to seek help, knowledge about and attitudes towards help-seeking as a key component of mental health literacy.

However, while there have been numerous mental health literacy interventions applied and evaluated in the past two decades [27-28, 32], there has been limited research determining the quality of measurement tools that evaluate mental health literacy interventions. There have been a few literature reviews describing mental health literacy measurement tools [34-36], but all focused only on the stigma component of mental health literacy, and none included the evaluation of help-seeking measures. We recently conducted a scoping review of available mental health literacy measurement tools [39] to define the scope of measures in mental health literacy and identify the need to examine the quality of mental health literacy measurement tools. In this chapter, we conducted a systematic review to evaluate the quality of mental health help-seeking tools addressing help-seeking behaviours, help-seeking intentions, attitudes towards help-seeking, and

[5] This chapter is accepted by the Journal of Mental Health [42].

knowledge about help-seeking among all population groups. The quality of mental health literacy tools addressing mental health knowledge and stigma are reported in Chapter 3 and 5 respectively.

5. 2 METHODS

We applied the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) manual for the critical appraisal of included studies [214] as COSMIN is the consensus based tool applied in systematic reviews to assess the quality of psychometric studies. We further evaluated the quality of each included measurement property, using criteria created by the COSMIN group [205]. We reported findings of this review based on the protocol recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [230].

5. 2. 1 Search Strategy

We conducted the search between January and June 2015 and updated the search between March and April of 2016. The search strategy covered three outcomes of mental health literacy (knowledge, stigma, and help-seeking), identified by the scoping review [39], created and applied in consultation with a health librarian. We included the results of the other two outcomes (knowledge and stigma) of mental health literacy until the stage of data extraction because mental health literacy outcomes can be integrated in one tool. We searched PubMed, PsycINFO, EMBASE, CINAHL, the Cochrane Library, and ERIC databases for relevant studies, using four sets of key words and phrases from the scoping review [39], with one set particularly addressing help-seeking. This includes terms addressing: general mental health and mental disorders; 3 outcomes of mental health literacy; assessment tools; and study designs. The search results of these 4 sets of terms were linked using the Boolean operator "AND". We also searched Google Scholar

using the names of relevant tools from the electronic database search. We further checked the references of included studies and searched Google Scholar for additional help-seeking tools. The search strategies used to search PubMed is presented in appendix 1.

Two reviewers independently searched the databases, screened studies (titles, abstract, full text scanning and reviewing) for inclusion and extracted the data. They met at the last stage of data screening to share their results and reached consensus on their final inclusion of studies. Differences in their results were discussed and consensus was reached on the final findings presented in this review. A group of mental health professionals and methodologists were on the team to advise on the differences of findings between the two reviewers.

5. 2. 2 Selection Criteria

We included studies using any study design that described and evaluated any measurement properties (reliability, validity, and responsiveness) of help-seeking tools. We only focused on help-seeking tools addressing mental health in general as well as tools on 4 common mental disorders: anxiety, depression, attention deficit hyperactivity disorder (ADHD), and schizophrenia. Age of participants or publication dates was not restricted by us. We included studies published in English.

We excluded studies that only provided the psychometrics of the properties but not descriptions of statistical analysis. For example, there were many studies evaluating mental health literacy interventions that only reported the internal consistency (e.g. Chronbach's alpha) of the tool used but failed to mention how it was obtained. As this review only addresses the help-seeking tools, knowledge and stigma tools were not included in the analysis.

5. 2. 3 Data Extraction

We extracted basic information about study characteristics, such as population, study type, location, year of publication, and the content of each tool included. We also collected data on measurement properties in the studies. We developed a data extraction form a priori based on the recommendations of the COSMIN checklist manual [214] and documented information on 9 measurement properties: 1. reliability (internal consistency, reliability (test-retest and intra-rater reliability), measurement errors); 2. validity (content validity, structural validity (factor analysis), hypothesis testing (construct validity), cross-cultural validity, and criterion validity); and 3. responsiveness (sensitivity to change). In data extraction, we considered tools that took or adapted items from other tools as independent from the original ones.

5. 2. 4 Study Quality Assessment (Risk of Bias)

We determined the quality of a study for each measurement property by the COSMIN criteria [214] that were applied to the 9 measurement properties described above. For each property, COSMIN checklist developed 7-18 criteria items to be evaluated and each item was ranked as “excellent”, “good”, “fair”, or “poor”. Then the final quality of a study on a measurement property was ranked as “excellent”, “good”, “fair”, and “poor”, based on the “worse score counts” approach. This means the methodological quality score of a study on a particular property was obtained by taking the lowest rating of any item out of the 7-18 items of criteria (‘worse score counts’). For example, if a study on the construct validity of a tool is assessed, and it has a range of rankings from “poor” to “excellent” on different items, then in this case, “poor” is the ranking for the final quality of the study on the construct validity as it is the worst score amongst the items ranked.

5. 2. 5 Quality of Measurement Properties and Levels of Evidence of Overall Quality

We further determined the level of evidence of the overall quality of each measurement property, considering the methodological quality of studies as described above and the consistency of the quality of each measurement property. We first rated the quality of each measurement property, using the evidence synthesis approach developed by Terwee and colleagues [205], which was refined in a systematic review by Uijen and colleagues [215] and consequently, each property was rated as: positive (+, above the quality threshold), negative (-, below the quality threshold), or indeterminate (?, lack of information to determine the quality). The quality criteria can be found in Appendix 2.

Following this, we determined the level of evidence on overall quality of a measurement property, considering the study quality (excellent, good, fair and poor) and its consistency, and the quality of each measurement property (+, -, ?) together. Our synthesis of the overall evidence was informed by the work of Uijen and colleagues [215], and by the recommendations on overall level of evidence by the Cochrane Back and Neck Group [216]. As a result, the level of evidence was ranked as strong (+++, or ---), moderate (++, or --), limited (+ or -), conflicting (+/-) or unknown (x) (Appendix 3). In this approach, the rating "x" originally referred to only studies of poor methodological quality, independent of the result of the measurement property [215]. However, this approach missed the possibility of the rated measurement properties "?" in studies of "fair", "good" or "excellent" methodological quality. In such cases, we rated the overall evidence as "unknown" (x) as well, since the quality of the measurement properties were unknown, despite the quality of studies with which they were evaluated.

We considered measurement properties with positive strong evidence (+++) as "ideal", moderate positive evidence (++) as "preferred", and limited positive evidence (+)

as “minimum acceptable”. We recommended the application of tools with such ratings based on the current available evidence.

Two reviewers independently rated the quality of included studies and related measurement properties, and they met at the last stage to discuss and reach consensus on their final ratings. Two mental health experts and one expert systematic research methodologist were available for consultation to assist with resolving any differences the two reviewers had for their final ratings.

5. 3 RESULTS

5. 3. 1 Study Selection

Figure 1 demonstrates the result of the search strategy employed. We imported data into Reference 2.0 database management software [46] and removed duplicates. We then screened 21089 studies, and excluded studies that were not the topic of interest at this stage (e.g., studies addressing HIV/AIDS stigma, Cognitive Behavioral Therapy, resilience, social and emotional learning, mental disorders that were not the topic of interest of this review). We repeated this procedure to further screened out irrelevant studies until the last stage when we excluded studies measuring mental health knowledge or stigma or mental illness. Google Scholar search followed the last stage of screening, and these processes together resulted in 24 studies that reported and analyzed psychometric properties of 12 help-seeking measurement tools [54, 120, 162-178, 219, 253-256]. Table 1 described the names and abbreviations of each tool, and we used tool acronyms in the discussion for the ease of reading. One tool, IASMHS [171] was created based on the ATHSS [162]; it however has made a number of adaptations (10 items) to the original one and has also added two additional factors. Therefore, we considered it as a separate tool.

5. 3. 2 Study Characteristics

Table 1 presents the detailed study characteristics of included studies. Studies were conducted among post-secondary school students (15 studies), high school students or youth in community (6 studies), community members (3 studies), elite athletes (1 study), patients (1 study), Christians (1 study), and caregivers (1 study). Studies took place mostly in developed countries such as the United States (11 studies), Australia (7 studies), Canada (1 study), Singapore (1study) except 2 studies conducted in Turkey, 1 in China (Taiwan), and 1 in Philippine.

We found that 7 tools (ATHSS, ATSPPH, ASPH-S, HSA, HSAS, PATPSI, IASMHS) [54, 120, 162, 165-167, 171-172, 174-177, 253, 255] evaluated attitudes and beliefs towards psychological help-seeking. They focused on evaluating the perceived need for seeking psychological help, level of trust with mental health professionals and concerns about stigma against seeking help. In addition, one tool, Jorm MHL [168-169, 173] on help seeking evaluated attitudes and beliefs towards interventions and help-seeking in general by listing a range of help-seeking resources for respondents to rate their helpfulness or harmfulness for a particular mental disorder, usually Depression or Anxiety Disorders. It listed a total of 35 help-seeking resources for respondents to rate, such as interventions by mental health professionals (e.g., family physicians, counselors, psychologists, psychiatrists, social workers); seeking help from families and friends; taking alternative health products such as vitamins or herbs; self-help strategies such as exercise, religions, reading; and consulting websites. One tool, MHLS [219] evaluated both attitudes towards and knowledge about help-seeking. The remaining 3 tools, ISCI, GHSQ, HSI [54, 163-164, 170, 178, 254, 256], all evaluated self-reported intentions to seek help. ISCI addressed the likelihood of students seeking counseling service for a

particular problem, while GHSQ and HSI evaluate the likelihood of seeking help from a variety of resources. No tools evaluating actual help-seeking behaviors for mental health were identified in our search (or included in our review).

5. 3. 3 Methodological Quality of Studies

Table 2 presents the methodological quality of studies regarding measurement properties of a specific tool. The 12 tools evaluated the following properties: internal consistency, reliability, content validity, structural validity (factor analysis), measurement errors, criterion validity and construct validity (hypothesis testing). Measurement properties of three tools, ATHSS (internal consistency, reliability, structural validity, construct validity) [162, 167, 174, 255], ATSPPH (internal consistency, reliability, structural validity, construct validity) [54, 120, 165-166, 253], and GHSQ (internal consistency, reliability, structural validity, construct validity) [54, 164, 178, 254, 256] were evaluated in multiple studies. The psychometrics of the remaining tools were evaluated only in one study each. No studies have been identified to address responsiveness (sensitivity to change) or cross-cultural validity although one study on ATHSS-Chinese discussed cultural adaptation of tools used [253].

The quality of included studies ranges from “poor” to “excellent” with 4 studies with “excellent” quality: one on the structural validity of the Jorm MHL tool [169]; two on the content validity of ATHSS [162] and IASMHS [171]; and one on the internal consistency, content validity and structural validity of [219]. The quality of studies on the same measurement property of a specific tool in different studies also varies, such as studies on ATHSS [162, 167, 174, 255] measuring its internal consistency (“poor” or “good”), reliability (“poor” or “fair”), content validity (“poor” or “excellent”), and structural validity (“fair” or “good”); and studies on GHSQ [54, 164, 178, 254, 256]

measuring its internal consistency (“poor” or “good”), and reliability (“poor” or “fair”). The quality of studies on 9 tools (ATSPPH, HSAS, HSA, GHSQ, ISCI, Jorm MHL, PATPSI, IASMHS, & MHLS) [120, 163, 165-166, 171-173, 176-178, 219, 256] measuring the construct validity (hypothesis testing) was consistently ranked as “fair”.

5. 3. 4 Level of Evidence on the Overall Quality of Measurement Properties

The overall level of evidence was determined by the study quality (Table 2) and the quality of each measurement property (Table 3) based on quality criteria described in Appendix 3. Positive and negative ratings of each measurement property indicated whether the quality of a property reached above the threshold defined by the COSMIN group. Table 3 showed that the construct validity of all included tools met the quality criteria and the rest of properties showed mixed ratings for different tools.

The overall level of evidence for each measurement property of a specific tool across studies was reported in Table 4. Although overall quality is usually determined by multiple studies, most of measurement properties in this review were assessed in only one study on which their final ratings were based except the properties of ATHSS [162, 167, 174, 255], ATSPPH [54, 120, 165-166, 253] and GHSQ [54, 164, 178, 254, 256] which were assessed in more than 1 study.

Findings showed that all tools demonstrate mixed levels of evidence (+++/---, ++/--, +, -, +/-, or x) for their related measurement properties. Overall, 11 tools showed limited or above level of evidence for some of the assessed properties. Of these 11, 4 tools exhibited strong level of evidence: IASMHS (+++) [171], ATHSS (+++) [162], and MHLS [219] with their content validity, and Jorm MHL (---) [169] with its structural validity. A moderate level of evidence was found in the internal consistency (++) of ATSPPH, HSA, ISCI, HSI, PATPSI, ASPH-S, IASMHS [163, 165-166, 170-171, 175-

177]; the structural validity of IASMHS (--) and GHSQ (++) [164, 171, 256]; and the construct validity of ATHSS (++), ATSPPH (++), and GHSQ (++) [120, 162, 165-167, 178, 255-256]. Further a number of properties in some tools have limited evidence (+) for quality. These are: the reliability of ASPH-S [175], ATHSS [162, 255], GHSQ [178, 254, 256]; the structural validity of ATHSS, HSA, ISCI, ASPH-S, and PATPSI [163, 174-177, 255]; the construct validity of HSAS, ISCI, Jorm MHL, IASMHS, MHLS, HSA, and PATPSI [163, 171-173, 176-177, 219]; and the criterion validity of ATSPPH [166].

Conflicting evidence (+/-) was identified in the reliability of PATPSI and Jorm MHL [168, 177]. Finally, a number of tools showed unknown evidence (x) across various properties. These were: the internal consistency of HSAS [172] and GHSQ [54, 178, 254]; the reliability of IASMHS [171] and ATSPPH [54, 166]; the content validity of ATSPPH, GHSQ, and HSI [165, 170, 178]; the structural validity of ATSPPH, HSI, and MHLS [165-166, 219, 251]. Table 4 demonstrated more details of the overall levels of evidence.

As we discussed earlier, the levels of evidence of measurement properties varied for a number of factors. For example, the "poor" study quality shown in table 2, or the "indeterminate" (?) quality of measurement properties in table 3 impacted the overall quality of a measurement property and could reduce the property quality down to unknown (x) level. In addition, the conflicting evidence (+/-) was due mainly to inconsistent findings. Studies with strong or moderate negative evidence (--- or --) implied measurement properties were under quality threshold despite that study qualities may be "excellent" or "fair".

We recommend the application of tools with measurement properties rated as positive strong (ideal) (+++) or moderate (preferred (++), and the application of tools with limited evidence (minimum acceptable) (+) with caution. Table 4 provides details of such tools. Measurement tools of which all assessed properties reached to the level of limited (+) evidence or above include ISCI and ASPH-S [163, 175]. And the rest of the 10 tools with mixed ratings from unknown (x) to strong evidence (+++ or ---). Properties with negative (---, --, or -), conflicting (+/-) ratings, or unknown level of evidence (x) have yet to be investigated further before application.

5. 4 DISCUSSION

This systematic review included 24 studies that contained 12 help-seeking measurement tools assessing help-seeking intentions, knowledge about help-seeking resources, attitudes towards psychological help-seeking, and attitudes towards help-seeking in general. We provided detailed study characteristics (Table 1), and applied the COSMIN criteria to systematically assess the study quality for each individual measurement property and further synthesized the overall level of evidence of the included measurement properties. To our knowledge, this is the first review that systematically appraised the quality of psychometric studies and the related measurement properties of mental health help-seeking measurement tools.

The study characteristics of included studies demonstrate that studies have been mostly conducted in more developed countries and regions (n=21) and among young populations (n=20). This, however, may have limited the generalizability of the tools to other populations and settings. It is known that mental health care help-seeking may differ substantially across cultures, age, socio-economic status, sex or geographical location [257-258] and from the available literature it is not known if an instrument that

demonstrates acceptable measurement properties in one research situation can be applied to others.

We also noticed that the most evaluated property amongst the tools was internal consistency, while measurement errors, cross-cultural validity and responsiveness, were not or rarely assessed. The COSMIN checklist [214] does not provide specific guidance on what/which specific measurement properties should be assessed, nor does it specify how many properties should be evaluated to determine the quality and adequacy of a measurement tool. Based on our findings, we decided that only assessing internal or consistency or reliability may not be sufficient to be an adequate evaluation of the instrument. Key components should always be addressed, including: internal consistency, test-retest reliability (stability), construct validity (construct testing), and structure validity (factors of the measure). When the tool is being applied in different populations from those in which it was developed, cultural validity should also be addressed.

It is noted that none of the 12 tools measured actual help seeking behaviors. Tools mainly focused on help-seeking attitudes and intentions which may be significant predictors of actual help-seeking behaviors [259]. However, our review didn't include studies analyzing studies on measures of help-seeking behaviors not because there are not such measures existing. Instead, we are aware that a number of tools have been developed to measure actual help-seeking behaviors for mental health [17, 196, 202-203, 260] but none have been validated, and therefore they may be considered as the target of future research for validation.

In this review, methodological quality for most psychometric properties ranged from "poor" to "excellent" except for construct validity that was consistently rated as

"fair" across different tools and studies. This varied quality of studies was attributed to a number of shortcomings of the research design defined by the COSMIN group [205, 214]. In order to improve the study quality, we propose that a study evaluating the psychometrics of a measurement tool address basic design requirements to discuss the missing items and adequacy of the sample size. In addition, studies on the internal consistency have to assess the dimensionality (factor analysis) of the tool. Further, to assess the content validity of a measure, its items have to be reviewed by the target population for relevancy and purposes of the measure and its content domains have to be specified in advance. Lastly, to establish the construct validity of a measure, the hypotheses tested have to be predetermined to avoid the potential biased interpretation of the results by researchers.

We found when a measurement property (e.g., internal consistency, reliability and content validity) was evaluated in different studies, the results tended to be mixed and conflicting (+/-), implying instability/lack of replicability of the tool (Table 4). As most of the measurement properties were only evaluated in one study, future research may focus on how tools are applied in various populations, across different cultures and in different settings for their generalizability and stability.

Based on the current available evidence, we are confident about the application of tools with measurement properties of strong (ideal) or moderate (preferred) level of evidence with positive ratings (+++ or ++) and we suggest the application of tools with measurement properties of limited positive evidence (+) with caution. We don't recommend the application of tools with other levels of evidence until further research. However, the validation of measurement properties is an ongoing process and the levels

of evidence may vary as further studies are conducted in different settings with different populations.

In this review, we identified 2 measurement tools, ISCI and ASPH-S [163, 175] in all of which their measurement properties reached limited or above level of evidence. However, this doesn't necessarily mean that they are better tools than others because each tool measures different aspects of help-seeking and different measurement properties. For example, MHLS [219] measured knowledge and attitudes towards help-seeking with 6 related measurement properties rated as limited, moderate or excellent levels of evidence except for the measurement error as unknown. Another tool, ISCI [163] measured intentions to seek help with its all 3 evaluated measurement properties rated as either limited or moderate level of evidence. In this case, we can't decide that ISCI is better than MHLS. Instead, what is more important for tool selection is to investigate the evidence of each individual property and the relevance of the tool to practice.

5. 5 LIMITATIONS

This review has a number of limitations. First, the COSMIN checklist [205, 214] was originally developed as the quality criteria for the measurement properties of health status questionnaires. Some of the criteria may not best fit for help-seeking tools for mental health. For example, items 7 and 9 for the reliability property discussed the stability of patient health status and test conditions between two tests, which were not applicable to help-seeking tools, and therefore we have excluded them in our analysis. Second, we may have missed some eligible studies due to the fact that we included studies published in English only, and we did not check the grey literature or contact people who developed the included tools.

5. 6 CONCLUSIONS

To our knowledge, this systematic review is the first of its kind to provide a comprehensive and critical synthesis of the quality of studies on current available help-seeking measurement tools for mental health and the overall level of evidence of each included measurement property. It has thus addressed a considerable gap in the field that could inform future research for both researchers and practitioners. The results of our analysis may also help researchers or decision makers to determine which tools may be most appropriate for use given the current evidence. At this stage, we are not comfortable recommending one tool over another, but instead ask researchers to consider the evaluations of each tool as reported in Table 2, 3 and 4, and to also consider the population in which the tool will be applied. Decisions can then be made resulting from that consideration. At this time, no tool reviewed has demonstrated a substantial degree of generalizability across diverse populations.

CHAPTER 6: CONCLUSIONS

The scoping and systematic reviews included in this thesis identified 16 mental health knowledge tools, 12 help-seeking tools, and 101 stigma measurement tools. We provided comprehensive descriptions of the study characteristics of all included studies, conducted critical analysis of the methodological study quality, the quality of individual measurement properties, and the overall evidence of the measurement properties of the included tools, applying a standardized method, the COSMIN checklist and recommendations by the Cochrane Back and Neck review group on the level of evidence of each assessed outcome [205, 214, 216]. Taken together, this research is the first of its kind known to the author to investigate the quality of mental health literacy tools in the field.

This comprehensive and critical synthesis of current evidence has contributed to the body of knowledge related to mental health literacy as the first such research in the field by creating a compendium of current mental health literacy measurement tools and determining the quality of these tools. Furthermore, this thesis work has advanced knowledge regarding how to determine the level of evidence of mental health literacy measurement tools by creating comprehensive criteria based on past research for future researchers to apply when deciding the overall quality of mental health literacy measurement tools.

One major contribution of this research lies in the fact that this is the first time to aggregate measurement tools as guided within the most recent mental health literacy framework (23) in the field. Past research on mental health literacy have weighed more on the creation and impact of various interventions focusing on knowledge about and

symptoms and self-care strategies. The current research has pointed out the importance of the quality of mental health literacy measures for the first time, and further stressed the necessity to develop mental health literacy tools encompassing all related components: knowledge, stigma and help-seeking. And therefore it has the potential to become the guidance for future mental health literacy research, including the development of mental health literacy interventions and related measures. More specifically, tables in chapters 2-5 have provided easy and comprehensive access to detailed tool information, including the content of each tool, target population, evaluation site, measurement properties assessed and their quality. These tables can serve as "go-to" resources for community stakeholders or researchers interested in this field.

Consistent across the scoping review and 3 systematic reviews described in the current research, there have been disproportionally more validated stigma measurement tools than tools addressing mental health knowledge or help-seeking. Further, this research indicated that currently, there is a lack of measurement tools that integrated all the components of mental health literacy within one tool. Particularly noticeable to us was the need to include stigma measures into mental health literacy tools as the field, consistent with the evolution of health literacy, is developing towards a broader construct of mental health literacy that integrates stigma with knowledge and help-seeking [19, 20, 23, 24, 26]. Inclusion of stigma components in mental health literacy evaluation tools should now become necessary and essential as research [19, 23, 227-229] shows that changes in knowledge and stigma are related and together they make significant impact on mental health help-seeking intentions and behaviors. Therefore, future research may focus on the creation and validation of such mental health literacy tools. This

consideration in turn will help to guide and shape how mental health literacy interventions should be developed and implemented, especially in the school setting.

The recommendations made herein, pertaining to the use of specific mental health literacy tools with strong (ideal), moderate (preferred), or limited (minimum acceptable) level of evidence in future research provide researchers and practitioners with information that can be used in further advancing the study of mental health literacy. More specifically, this work has the potential to improve future study of mental health literacy by promoting researcher's use of measures that have demonstrated an appropriate degree of technical sophistication instead of applying measures that are less robust in their quality. Further, this thesis work has improved the understanding of what mental health literacy interventions should entail to address mental health knowledge, reduce stigma and encourage help-seeking behaviors. This information may be particularly useful in guiding the community, such as schools, to develop interventions to appropriately address the community mental health needs. It has thus addressed a considerable gap in the field.

Additionally, however, this research has also pointed out that the application of specific tools has to depend on the context in which they were developed and validated. For example, a well-validated measurement property in one study may not demonstrate similar robust outcome if it is applied in another location or different cultural context. Therefore, future research should focus both on improvements to current tools and the validation of these tools as they are applied in different contexts with different populations. More informed decisions about which tools are most appropriate in what circumstances can then be made. This is necessary as the validation of measurement

properties of mental health literacy tools is not a static event but rather an ongoing and iterative process [223].

Informed by the findings from these 3 systematic reviews (chapters 3, 4 and 5), we have developed and suggested criteria for tool choice. We have noted that a measurement tool should: have its items appropriately related (internal consistency); be stable over time (test-retest reliability); and have adequately established constructs and factors (structural and construct validity). Further cross-cultural validity has to be determined prior to its application in culturally and socially different settings. While we have comfort with these criteria we understand that other researchers in this field may provide considered comment on our recommendations that may result in our further refinement of these criteria. Should this occur, we would welcome the dialogue, as this would demonstrate a process of researcher interaction that to our knowledge has not yet occurred in the field of mental health literacy research.

This research has advanced the previous work to determine the level of evidence of overall quality of measurement properties across studies. Terwee and colleagues [205, 214] provided criteria for methodological quality of a psychometric study and criteria for the quality of each individual property in a single study. Uijen and colleagues (2013) [215] furthered the work by integrating and adapting criteria from the Cochrane Back and Neck group [216] to determine the level of evidence if a measurement property is assessed in multiple studies, however did not define how consistent the study findings should be in order to be assigned to a certain level of ranking. Our research created concrete steps to complete this process for researchers to follow and make further consensus on.

To summarize, currently available mental health literacy measurement tools have varied qualities due to technical quality challenges. We recommended the application of mental health literacy tools by the quality of each assessed measurement property because of the quality variation of these properties even in the same study of the tool. This indicates the strong need for researchers to follow guidelines, such as the COSMIN checklist [205, 214] to appropriately conduct psychometrics studies. Further, there is a need to develop and evaluate mental health literacy tools that embody the current understanding of the concept of mental health literacy, which in turn will guide the development of mental health literacy interventions.

REFERENCES

1. Costello EJ, Mustillo S, Keeler G, & Angold A. Prevalence of psychiatric disorders in childhood and adolescence. In: Levin BL, Petrila J, Hennessy K, editors. Mental health services: a public health perspective (2nd ed.). New York: Oxford University Press; 2004. p. 111-128.

2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR. Lifetime prevalence and age of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Archives of General Psychiatry. 2005; 62(6): 593-602.

3. Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry. 2007; 6(3): 168-176.

4. Whiteford HA1, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, Johns N, Burstein R, Murray CJ, Vos T. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet. 2013; 382(9904):1575-86.

5. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015 Mar;56(3):345-65. doi:10.1111/jcpp.12381.

6. World Health Organization. (2011) Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. 2011. Available from: http://apps.who.int/gb/ebwha/pdf_files/EB130/B130_9-en.pdf. [Accessed 10th June, 2016].

7. Bloom DE, Cafiero ET, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, Feigl AB, Gaziano T, Mowafi M, Pandya A, Prettner K, Rosenberg L, Seligman B, Stein AZ, &Weinstein C. The global economic burden of noncommunicable diseases. Geneva: World Economic Forum, 2011. Available from: http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf. [Accessed 10th June, 2016].

8. Bhatia S. Childhood and adolescent depression. American Family Physician. 2007; 75(1): 73.

9. Kessler RC, Foster C L, Saunders WB, Stang PE. Social consequences of psychiatric disorders I: educational attainment. The American Journal of Psychiatry. 1995;152(7): 1026-32.

10. Mezuk B, Eaton WW, Albrecht S, Golden, SH. Depression and type 2 diabetes over the lifespan. Diabetes Care. 2008; 31(12): 2383-90.

11. Rugulies R. Depression as a predictor for coronary heart disease: a review and meta-analysis. American Journal of Preventive Medicine. 2002; 23(1): 51-61.

12. Scott KM, Lim C, Al-Hamzawi A, Alonso J, Bruffaerts R, Caldas-de-Almeida JM, Florescu S, de Girolamo G, Hu C, de Jonge P, Kawakami N, Medina-Mora ME, Moskalewicz J, Navarro-Mateu F, O'Neill S, Piazza M, Posada-Villa J, Torres Y, Kessler RC. Association of mental disorders with subsequent chronic physical conditions: world mental health surveys from 17 countries. JAMA Psychiatry. 2016; 73(2): 150-158. doi: 10.1001/jamapsychiatry.2015.2688.

13. Leaf PJ, Alegria M, Cohen P, Goodman SH, Horwitz SM, Hoven CW, et al. Mental health service use in the community and schools: results from the four-community MECA study. J AM Acad Child Adolesc Psychiatry. 1996; 35: 889-97.

14. Thornicroft G. Most people with mental illness are not treated. Lancet. 2007; 370(9590): 807-8.

15. Waddell C, McEwan K, Shepherd CA, Offord D R, Hua JM. A public health strategy to improve the mental health of Canadian children. Can J Psychiatry. 2005; 50(4): 226-33.

16. Patel V, Flisher AJ, Hetrick S, et al. Mental health of young people: a global public-health challenge. Lancet. 2007; 369(9569):1302–1313.

17. Rickwood, D., Cavanagh, S., Curtis, L., & Sakrouge, R. Educating young people about mental health and mental illness: evaluating a school-based programme. International Journal of Mental Health Promotion. 2004; 6, 23–32.

18. Sharp W, Hargrove DS, Johnson L, Deal WP. Mental health education: an evaluation of a classroom based strategy to modify help seeking for mental health problems. Journal of College Student Development. 2006; 47 (4), 419-438.

19. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry. 2010;10: 113.

20. Rusch N, Evans-Lacko S, Henderson C, Flach C, Thornicroft G. Public knowledge and attitudes as predictors of help seeking and disclosure in mental illness. Psychiatric Services. 2011; 62(6): 675-8.

21. Corrigan PW, Watson AC. Factors that explain how policy makers distribute resources to mental health services. Psychiatric Services. 2003; 54(4): 501-7.

22. Henderson C, Evans-Lacko S, Thornicroft G. Mental illness stigma, help seeking, and public health programs. American Journal of Public Health. 2013; 103(5): 777-80. doi:10.2105/AJPH.2012.301056

23. Kutcher S, Wei Y, Coniglio C. Mental health literacy: past, present, and future. Canadian Journal of Psychiatry. 2016; 61(3): 154-158.

24. World Health Organization. The solid facts: health literacy. World Health Organization, 2013. Available from: http://www.euro.who.int/__data/assets/pdf_file/0008/190655/e96854.pdf. [Accessed 15th August, 2016]

25. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. The Medical Journal of Australia. 1997; 166(4): 182-6.

26. Kutcher S, Bagnell A, Wei Y. Mental health literacy in secondary schools: a Canadian approach. Child and Adolescent Psychiatric Clinics of North America. 2015; 24(2): 233-44.

27. Wei Y, Hayden JA, Kutcher S, Zygmunt A, McGrath P. The effectiveness of school mental health literacy programs to address knowledge, attitudes and help seeking among youth. Early Intervention Psychiatry. 2013; 7(2): 109-21.

28. Kelly CM, Jorm AF, Wright A. Improving mental health literacy as a strategy to facilitate early intervention for mental disorders. Medical Journal of Austria. 2007;187 Suppl 7: 26-30.

29. Schachter HM, Girardi A, Ly M, Lacroix D, Lumb AB, van Berkom J, Gill R. Effects of school-based interventions on mental health stigmatization: a systematic review. Child Adolesc Psychiatry Ment Health. 2008; 2(1):18. doi: 10.1186/1753-2000-2-18.

30. Angermeyer MC, Dietrich S. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatr Scand. 2006; 113(3): 163-79.

31. Mittal D, Sullivan G, Chekuri L, Allee E, Corrigan PW. Empirical studies of self-stigma reduction strategies: a critical review of the literature. Psychiatr Serv. 2012; 63(10): 974-81.

32. Hadlaczky G, Hokby S, Mkrtchian A, Carli V, Wasserman D. Mental health first aid is an effective public health intervention for improving knowledge, attitudes, and behaviors: a meta-analysis. Int Rev Psychiatry. 2014; 26(4): 467-475. doi: 10.3109/09540261.2014.924910.

33. Weare K, & Nind M. Mental health promotion and problem prevention in schools: what does the evidence say? Health Promot Int. 2011; Suppl 1: i29-69. doi:10.1093/heapro/dar075.

34. Link BG, Yang LH, Phelan JC, Collins, PY. Measuring mental illness stigma. Schizophr Bull. 2004; 30(3): 511-41.

35. Boyd JE, Adler EP, Otilingam PG, Peters T. Internalized Stigma of Mental Illness (ISMI) scale: a multinational review. Compr Psychiatry.2014; 55(1): 221-31. doi: 10.1016/j.comppsych.2013.06.005.

36. Brohan E, Slade M, Clement S, Thornicroft G (2010) Experiences of mental illness stigma, prejudice and discrimination: a review of measures. BMC Health Services Research. 2010; 10: 80. Available from: http://www.biomedcentral.com/1472-6963/10/80. [Accessed 15th August, 2016].

37. O'Connor M, Casey L, Clough B. Measuring mental health literacy--a review of scale-based measures. J Ment Health. 2014; 23(4):197-204.

38. Goffman E. Stigma: notes on the management of spoiled identity. Englewood Cliffs, NJ: Prentice Hall; 1963.

39. Wei Y, McGrath P, Hayden J, & Kutcher S. Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. BMC Psychiatry. 2015;15, 291. doi: 10.1186/s12888-015-0681-9

40. Wei Y, McGrath P, Hayden J, & Kutcher S. Measurement properties of tools measuring mental health knowledge: a systematic review. BMC Psychiatry. 2016; 16 (1). doi: 10.1186/s12888-016-1012-5.

41. Wei Y, McGrath P, Hayden J, & Kutcher S. The Quality of Measurement Tools Evaluating the Stigma of Mental Illness: A Systematic Review. Epidemiology and Psychiatric Sciences. Submitted.

42. Wei Y, McGrath P, Hayden J, & Kutcher S. Measurement properties of mental health literacy tools measuring help-seeking: a systematic review. Journal of Mental Health. Accepted.

43. World Health Organization. WHO methods and data sources for global burden of disease estimates 2000-2011. World Health Organization. 2013. Available from: http://www.who.int/healthinfo/statistics/GlobalDALYmethods_2000_2011.pdf?ua=1. [Accessed 13 Aug 2015].

44. Reavley NJ, Jorm AF. National survey of mental health literacy and stigma. Canberra: Department of health and ageing. 2011. Available from: http://pmhg.unimelb.edu.au/research_settings/general_community/?a=636496. [Accessed 14 Aug 2015].

45. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int. J. Social Research Methodology. 2005; 8(1): 19-32.

46. RefWorks-COS PL, ProQuest LLC. RefWorks. 2nd ed. Ann Arbour, MI: ProQuest LLC; 2001.

47. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, R.G. Burgess, RG, editors. Analysing qualitative data. London: Routledge; 1994.

48. Ascher-Svanum H. Development and validation of a measure of patients' knowledge about schizophrenia. Psychiatr Serv. 1999; 50(4): 561-3.

49. Compton MT, Quintero L,Esterberg ML. Assessing knowledge of schizophrenia: development and psychometric properties of a brief, multiple-choice knowledge test for use across various samples. Psychiatry Research. 2007; 151: 87-95. doi: 10.1016/j.psychres.2006.05.019

50. Compton MT, Hankerson-Dyson D, Broussard B. Development, item analysis, and initial reliability and validity of a multiple-choice knowledge of mental illnesses test for lay samples. Psychiatry Research. 2011; 189: 141-8. doi: 10.1016/j.psychres.2011.05.041.

51. Evans-Lacko S, Little K, Meltzer H, Rose D, Rhydderch D, Henderson C, Thornicroft G. Development and psychometric properties of the mental health knowledge schedule. Can J of Psychiatry. 2010; 55(7): 440-8.

52. Gabriel A, Violato C. The development of a knowledge test of depression and its treatment for patients suffering from non-psychotic depression: a psychometric assessment. BMC Psychiatry. 2009; 9: 56. doi: 10.1186/1471-244X-9-56.

53. Kiropoulos L, Griffiths KM, Blashki G. Effects of a multilingual information website intervention on the levels of depression literacy and depression-related stigma in Greek-born and Italian-born immigrants living in Australia: a randomized controlled trial. J Med Internet Res. 2011; 13(2): e34. doi: 10.2196/jmir.1527.

54. Gulliver A, Griffiths KM, Christensen H, Mackinnon A, Calear AL, Parsons A, Bennett K, Batterham PJ, Stanimirovic R.Internet-based interventions to promote mental health help-seeking in elite athletes: an exploratory randomized controlled trial. J Med Internet Res. 2012;14(3): e69. doi: 10.2196/jmir.1864.

55. Hepperlen TM, Clay DL, Henly GA, Barké CR, Hehperlen MH, Clay DL. Measuring teacher attitudes and expectations toward students with ADHD: development of the test of knowledge about ADHD (KADD). J Atten Disord.2002; 5(3): 133-42. doi: 10.1177/108795470200500301.

56. Kronmüller KT, Saha R,Kratz B, Karr M, Hunt A, Mundt C, Backenstrass M. Reliability and validity of the Knowledge about Depression and Mania Inventory. Psychopathology. 2008; 41: 69-76. doi: 10.1159/000111550.

57. Pickett-Schenk SA, Cook JA, Laris A. Journey of hope program outcomes. Community Mental Health Journal. 2000; 36(4): 413-24.

58. Serra M, Lai A, Buizza C, Pioli R, Preti A, Masala C, Petretto DR. Beliefs and attitudes among Italian high school students toward people with severe mental disorders. Journal of Nervous and Mental Disease. 2013; 201(4): 311-8.

59. Shelley RH, Kastelic EA, Wilcox HC, Beaudry MB, Musei RJ, Heley K, et al. Achieving depression literacy: the adolescent depression knowledge questionnaire (ADKQ). School Mental Health. 2014; 6: 213-23. doi: 10.1007/s12310-014-9120-1

60. Swami V, Persaud R, Furnham A. The recognition of mental health disorders and its association with psychiatric skepticism, knowledge of psychiatry, and the big five personality factors: an investigation using the overclaiming technique. Soc Psychiatry Psychiatr Epedemiol. 2011; 46: 181-9. doi: 10.1007/s00127-010-0193-3.

61. Wang J, He Y, Jiang Q, Cai J, Wang W, Zeng Q, et al. Mental health literacy among residents in Shanghai. Shanghai Archives of Psychiatry. 2013; 25(4): 224-35.

62. Andersson HW, Bjørngaard JH, Silje Lill Kaspersen SL, Wang CEA, Skre I, Dahl T. The effects of individual factors and school environment on mental health and prejudiced attitudes among Norwegian adolescents. Soc Psychiat Epidemiol. 2010; 45: 569-577. doi:10.1007/s00127-009-0099-0.

63. Angermeyer MC, Matschinger H. The stigma of mental illness: effects of labeling on public attitudes towards people with mental disorder. Acta Psychiatr Scand. 2003;108: 304-9.

64. Aromaa E, Tolvanen A, Tuulari J, Wahlbeck K. Attitudes towards people with mental disorders: the psychometric characteristics of a Finnish questionnaire. Soc Psychiat Epidemiol. 2010; 45: 265–73. doi: 10.1007/s00127-009-0064-y.

65. Bagley, C., & King, M. Exploration of three stigma scales in 83 users of mental health services: implications for campaigns to reduce stigma. Journal of Mental Health. 2005; 14(4): 343-55. doi: 10.1080/09638230500195270.

66. Barney LJ, Griffiths KM, Christensen H, Jorm AF. The self-stigma of depression scale (SSDS): development and psychometric evaluation of a new instrument. International Journal of Methods in Psychiatric Research. 2010; 19(4): 243-54. doi: 10.1002/mpr.325.

67. Bell L, Long S, Garvan C, Bussing R. The impact of teacher credentials on ADHD stigma perceptions. Psychology in the Schools. 2011; 48(2): 184-97.doi: 10.1002/pits.20536.

68. Bogardus E. Measuring social distance. Journal of Applied Sociology. 1925; 9: 299-308.

69. Botega N, Mann A, Blizard R, Wilinson G. General practitioners and depression – first use of the depression attitude questionnaire. International Journal of Methods in Psychiatric Research. 1992; 2: 169-80.

70. Boyd JE, Otilingam PG. Brief version of the internalized stigma of mental illness (ISMI) scale: psychometric properties and relationship to depression, self esteem, recovery orientation, empowerment, and perceived devaluation and discrimination. Psychiatric Rehabilitation Journal. 2014; 37(1): 17-23. doi: 10.1037/prj000003517.

71. Brohan E, Clement S, Rose D, Sartorius N, Slade M, Thornicroft G. Development and psychometric evaluation of the discrimination and stigma scale (DISC). Psychiatric Research. 2013; 208: 33-40. doi: 10.1016/j.psychres.2013.03.007.

72. Brown SA. Factors and measurement of mental illness stigma: a psychometric examination of the attribution questionnaire. Psychiatric Rehabilitation Journal. 2008; 32(2): 89-94. doi:10.2975/32.2.2008.89.94.

73. Bjorkman T, Svensson B, Lundberg B. Experiences of stigma among people with severe mental illness. Reliability, acceptability and construct validity of the Swedish versions of two stigma scales measuring devaluation/discrimination and rejection experiences. Nord J Psychiatry. 2007; 61: 332-8. doi: 10.1080/08039480701642961.

74. Burra P, Kalin R, Leichner P, Waldron JJ, Handforth JR, Jarrett FJ, Amara IB. The ATP 30 – a scale for measuring medical students' attitudes of psychiatry. Medical Education. 1982; 16: 31-8.

75. Chang C, Wu T, Chen C, Wang J, Lin C. Psychometric evaluation of the internalized stigma of mental illness scale for patients with mental illnesses: measurement invariance across time. PLoS One. 2014; 9(6): e98767. doi: 10.1371/journal.pone.0098767.

76. Chowdhury AN, Sanyal D, Dutta SK, Banerjee S, De R, Bhattacharya K, et al. International Medical Journal. 2000; 7(4): 257-60.

77. Clayfield JC, Fletcher KE, Grudzinskas Jr, AJ. Development and validation of the mental health attitude survey for police. Community Ment Health J. 2011; 47: 742–51. doi: 10.1007/s10597-011-9384-y.

78. Cohen J, Struening EL. Opinions about mental illness in the personnel of two large mental hospitals. Journal of Abnormal and Social Psychology. 1962; 64(5): 349-60.

79. Corrigan PW, Rowan D, Qreen A, Lundin R, River P, Uphoff'-Wasowski K, et al. Challenging two mental illness stigmas: personal responsibility and dangerousness. Schizophrenia Bulletin. 2002; 28(2): 293-309.

80. Corrigan P, Markowitz FE, Watson A, Rowan D, Kubiak MA. An attribution model of public discrimination towards persons with mental illness. Journal of Health and Social Behavior. 2003; 44(2): 162-79.

81. Corrigan PW, Michaels PJ, Vega E, Gause M, Watson AC, Rusch N. Self-stigma of mental illness scale-short form: reliability and validity. Psychiatry Research. 2012; 199: 65-9. doi: 10.1016/j.psychres.2012.04.009.

82. Corrigan PW, Watson AC, Barr L. The self-stigma of mental illness: implications for self-esteem and self-efficacy. Journal of Social and Clinical Psychology. 2006; 25(9): 875-84.

83. Dalky HF. Arabic translation and cultural adaptation of the stigma-devaluation scale in Jordan. Journal of Mental Health. 2012; 21(1): 72–82. doi: 10.3109/09638237.2011.629238.

84. Day EN, Edgren K, Eshleman A. Measuring stigma toward mental illness: development and application of the mental illness stigma scale. Journal of Applied Social Psychology. 2007; 10: 2191-219.

85. Diksa E, Rogers ES. Employer concerns about hiring persons with psychiatric disability: results of the employer attitude questionnaire. Rehabilitation Counseling Bulletin. 1996; 40(1): 31-44.

86. Evans-Lacko S, Henderson C, Thornicroft G. Public knowledge, attitudes and behavior regarding people with mental illness in England 2009-2012. British Journal of Psychiatry, 2013; 202 Suppl: 51-7. doi: 10.1192/bjp.bp.112.112979.

87. Evans-Lacko S, London J, Japhet S, Rusch N, Flach C, Corker E, et al. Mass social contact interventions and their effect on mental health related stigma and intended discrimination. BMC Public Health. 2012; 12: 489.

88. Evans-Lacko S, Rose D, Little K, Flach C, Rhydderch D, Henderson C, Thornicroft G. Development and psychometric properties of the reported and intended behavior scale (RIBS): a stigma-related behavior measure. Epidemiology and Psychiatric Sciences. 2011; 20: 263-71.

89. Friedrich B, Evans-Lacko S, London J, Rhydderch D, Henderson C, Thornicroft G. Anti-stigma training for medical students: the education not discrimination project. British Journal of Psychiatry. 2013; 202 Suppl: 89-94. doi:10.1192/bjp.bp.112.114017.

90. Fuermaier ABM, Tucha L, Koerts J, Mueller AK, Lange KW, Tucha O. Measurement of stigmatization towards Adults with Attention Deficit Hyperactivity Disorder. PLoS ONE. 2012; 7(12): e51755. doi:10.1371/journal.pone.0051755.

91. Gabbidon J, Brohan E, Clement S, Henderson RC, Thornicroft G, MIRIAD Study Group. The development and validation of the Questionnaire on Anticipated Discrimination (QUAD). BMC Psychiatry.2013; 13: 297. Available from: http://www.biomedcentral.com/1471-244X/13/297. [Accessed 14th Aug 2015].

92. Gabbidon J, Clement S, Nieuwenhuizen A, Kassam A, Brohan E, Norman I, Thornicroft G. Mental Illness: Clinicians' Attitudes (MICA) Scale—Psychometric properties of a version for health care students and professionals. Psychiatry Research. 2013; 206: 81-7. doi: 10.1016/j.psychres.2012.09.028.

93. Gabriel A, Violato C. The development and psychometric assessment of an instrument to measure attitudes towards depression and its treatments in patients suffering from non-psychotic depression. Journal of Affective Disorders. 2010; 124: 241-9.doi:10.1016/j.jad.2009.11.009.

94. Gibbons C, Dubois S, Morris K, Parker B, Maxwell H, Bédard M. The development of a questionnaire to explore stigma from the perspective of individuals with serious mental illness. Canadian Journal of Community Mental Health. 2012; 31(1): 17-32.

95. Glozier N, Hough C, Henderson M, Holland-Elliott K. Attitudes of nursing staff towards co-workers returning from psychiatric and physical illnesses. International Journal of Social Psychiatry. 2006; 52(6): 525-34. doi: 10.1177/0020764006066843.

96. Granello DH, Pauley PS, Carmichael A. Relationship of the media to attitudes toward people with mental illness. The Journal of Humanistic Counseling, Education and Development. 1999; 38(2):98–110.

97. Granello DH, Pauley PS. Television viewing habits and their relationship to tolerance toward people with mental illness. Journal of Mental Health Counseling. 2000; 22(2): 162-75.

98. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatizing attitudes to depression: randomized controlled trial. British Journal of Psychiatry. 2004; 185: 342-9. doi:10.1192/bjp.185.4.342.

99. Griffiths KM, Christensen H, Jorm AF. Predicators of depression stigma. BMC Psychiatry. 2008; 8: 25.doi:10.1186/1471-244X-8-25.

100. Griffiths, KM, Batterham PJ, Barney L, Parsons A. The generalized anxiety stigma scale (GASS): psychometric properties in a community sample. BMC Psychiatry. 2011; 11: 184. Available from: http://www.biomedcentral.com/1471-244X/11/184. [Accessed 14th Aug 2015].

101. Haddad M, Menchetti M, McKeown E, Tylee A, Mann A. The development and psychometric properties of a measure of clinicians' attitudes to depression: the revised Depression Attitude Questionnaire (R-DAQ). BMC Psychiatry. 2015; 15: 7. doi: 10.1186/s12888-014-0381-x.

102. Harvey RD. Individual differences in the phenomenological impact of social stigma. The Journal of Social Psychology. 2001; 141(2): 174-89.

103. Hayward P, Wong G, Bright JA, Lam D. Stigma and self-esteem in manic depression: an exploratory study. Journal of Affective Disorders. 2002; 69: 61-7.

104. Hinkelmean L, Granello DH. Biological sex, adherence to traditional gender roles, and attitudes toward persons with mental illness: an exploratory investigation. Journal of Mental Health Counseling. 2003; 25(4): 259-70.

105. Hirai M, Clum GA. Development, reliability, and validity of the beliefs toward mental illness scale. Journal of Psychopathology and Behavioral Assessment. 2000; 22(3): 221-36.

106. Ho AHY, Potash JS, Fong TCT, Ho VFL, Chen EYH, Lau RHW, et al. Psychometric properties of a Chinese version of the stigma scale: examining the complex experience of stigma and its relationship with self-esteem and depression among people living with mental illness in Hong Kong. Comprehensive Psychiatry. 2015; 56: 198-205. doi:10.1016/j.comppsych.2014.09.016

107. Högberg T, Magnusson A, Ewertzon M, Lützén K. Attitudes towards mental illness in Sweden: adaptation and development of the Community Attitudes towards Mental Illness questionnaire. International Journal of Mental Health Nursing. 2008; 17: 302-10. doi: 10.1111/j.1447-0349.2008.00552.x.

108. Interian A, Ang A, Gara MA, Link B, Rodriguez MA, Vega WA. Stigma and depression treatment utilization among Latinos: utility of four stigma measures. Psychiatric Services. 2010; 61: 373-9.

109. Isaac F, Greenwood KM, Benedetto M. Evaluating the psychometric properties of the attitudes towards depression and its treatments scale in an Australian sample. Patient Preference and Adherence. 2012; 6: 349-54. doi:10.2147/PPA.S26783.

110. Jackson D, Heatherington L. Young Jamaicans' attitudes toward mental illness: experimental and demographic factors associated with social distance and stigmatizing opinions. Journal of Community Psychology. 2006; 34(5): 563-76. doi: 10.1002/jcop.20115.

111. Kanter JW, Rusch LC, Brondino MJ. Depression Self-Stigma: a new measure and preliminary findings. The Journal of Nervous and Mental Disease. 2008; 196 (9): 663-70. doi:10.1097/NMD.0b013e318183f8af.

112. Karidi MV, Vasilopoulou D, Savvidou E, Vitoratou S, Rabavilas AD, Stefanis CN. Aspects of perceived stigma: the stigma inventory for mental illness, its development, latent structure and psychometric properties. Comprehensive Psychiatry. 2014; 55: 1620-1625. doi:10.1016/j.comppsych.2014.04.002.

113. Kassam A, Glozier N, Leese M, Henderson C, Thornicroft G. Development and responsiveness of a scale to measure clinicians' attitudes to people with mental illness (medical student version). Acta Psychiatr Scand. 2010; 122: 153–61. doi: 10.1111/j.1600-0447.2010.01562.x.

114. Kassam A, Glozier N, Leese M, Loughran J, Thornicroft G. A controlled trial of mental illness related stigma training for medical students. BMC Medical Education. 2011; 11: 51. Available from: http://www.biomedcentral.com/1472-6920/11/51. [Accessed 14th Aug 2015].

115. Kira IA, Ramaswamy V, Lewandowski L, Mohanesh J, Abdul-Khalek H. Psychometric assessment of the Arabic version of the Internalized Stigma of Mental Illness (ISMI) measure in a refugee population. Transcult Psychiatry. 2015; 1-23. doi: 10.1177/1363461515569755.

116. Kassam A, Papish A, Modgill G, Patten S. The development and psychometric properties of a new scale to measure mental illness related stigma by health care providers: the opening minds scale for Health Care Providers (OMS-HC). BMC Psychiatry. 2012; 12: 62. http://www.biomedcentral.com/1471-244X/12/62. Accessed 14 Aug 2015.

117. Kellison I, Bussing R, Bell L, Garvan C. Assessment of stigma associated with attention-deficit hyperactivity disorder: psychometric evaluation of the ADHD stigma questionnaire. Psychiatry Research. 2010; 178: 363-9.doi:10.1016/j.psychres.2009.04.022

118. King M, Dinos S, Shaw J, Watson R, Stevens S, Passetti F, et al. (2007). The stigma scale: development of a standardized measure of the stigma of mental illness. British Journal of Psychiatry. 2007; 190: 248-54. doi: 10.1192/bjp.bp.106.024638.

119. Kobau, R, DiIorio C, Chapman D, Delvecchi P, SAMHSA/CDC Mental Illness Stigma Panel Members. Attitudes about mental illness and its treatment: validation of a generic scale for public health surveillance of mental illness associated stigma. Community Mental Health Journal. 2010; 46:164–76. doi: 10.1007/s10597-009-9191-x.

120. Komiya N, Good GE, Sherrod NB. Emotional openness as a predictor of college students' attitudes toward seeking psychological help. Journal of Counseling Psychology. 2000; 47(1): 138-43. doi: 10.1037//0022-0167.47.1.138.

121. Lam DCK, Salkovskis PM, Warwick HM. C. An experimental investigation of the impact of biological versus psychological explanations of the cause of "mental illness". Journal of Mental Health. 2005; 14(5): 453-64. doi: 10.1080/09638230500270842.

122. Lien Y, Kao Y, Liu Y, Chang H, Tzeng N, Lu C, Loh C. Internalized stigma and stigma resistance among patients with mental Illness in Han Chinese population. Psychiatr Q. 2014;86(2): 181-97. doi: 10.1007/s11126-014-9315-5.

123. Link BG. Reconsidering the social rejection of ex-mental patients: levels of attitudinal response. American Journal of Community Psychology.1983; 11(3): 261-73.

124. Link BG. Understanding labeling effects in the area of mental disorders: an assessment of the effects of expectations of rejection. American Sociological Review. 1987; 52(1): 96-112.

125. Link BG, Cullen FT, Frank J, Wozniak JF. The social rejection of former mental patients: understanding why labels matter. American Journal of Sociology.1987; 92(6): 1461-1500.

126. Luca Pingani L, Forghieri M, Ferrari S, Ben-Zeev D, Artoni P, Mazzi F, et al. Stigma and discrimination toward mental illness: translation and validation of the Italian version of the attribution questionnaire-27 (AQ-27-I). Soc Psychiatry Psychiatr Epidemiol. 2012; 47: 993–9. doi: 10.1007/s00127-011-0407-3.

127. Luty J, Fekadu D, Umoh O, Gallagher J. Validation of a short instrument to measure stigmatized attitudes towards mental illness. Psychiatric Bulletin. 2006; 30: 257-60. doi: 10.1192/pb.30.7.257.

128. Luty J, Umoh O, Sessay M, Sarkhel A. Effectiveness of Changing Minds campaign factsheets in reducing stigmatised attitudes towards mental illness. Psychiatric Bulletin. 2007; 31: 377-81. doi: 10.1192/pb.bp.106.012443.

129. Madianos M, Economou M, Peppou LE, Kallergis G, Rogakou E, Alevizopoulos G. Measuring public attitudes to severe mental illness in Greece: development of a new scale. Eur. J. Psychiat. 2012; 26(1): 55-67.

130. Magliano L, Marasco C, Guarneri M, Malangone C, Lacrimini G, Zanus P, Maj M. A new questionnaire assessing the opinions of the relatives of patients with schizophrenia on the causes and social consequences of the disorder: reliability and validity. Eur Psychiatry. 1999; 14: 71-5.

131. Magliano L, Fiorillo A, Rosa C, Malangone C, Maj M. Beliefs about schizophrenia in Italy: a comparative nationwide survey of the general public, mental health professionals, and patients' relatives. Canadian Journal of Psychiatry. 2004; 49: 322-30.

132. Mak WWS, Cheung RYM. Affiliate stigma among caregivers of people with intellectual disability or mental illness. Journal of Applied Research in Intellectual Disabilities. 2008; 21: 532-45. doi:10.1111/j.1468-3148.2008.00426.x.

133. Mansouri L, Dowell DA. Perceptions of stigma among the long-term mentally ill. Psychosocial Rehabilitation Journal. 1989; 13(1): 79-91.

134. McKeague L, Hennessy E, O'Driscoll C, Heary C. Peer Mental Health Stigmatization Scale: psychometric properties of a questionnaire for children and adolescents. Child and Adolescent Mental Health. 2015; doi: 10.1111/camh.12088.

135. Michaels PJ, Corrigan PW. Measuring mental illness stigma with diminished social desirability effects. Journal of Mental Health. 2013; 22(3): 218-226. doi:10.3109/09638237.2012.734652.

136. Minnebo J, Acker AV. Does television influence adolescents' perceptions of and attitudes toward people with mental illness? Journal of Community Psychology. 2004; 32(3): 257-75. doi: 10.1002/jcop.20001.

137. Modgill G, Knaak S, Kassam A, Szeto A. Opening minds stigma scale for health care providers (OMS-HC): examination of psychometric properties and responsiveness. BMC Psychiatry. 2014; 14:120. doi:10.1186/1471-244X-14-120.

138. Morris R, Scott PA, Cocoman A, Chambers M, Guise V, Va¨lima¨ki M, Clinton G. Is the Community Attitudes towards the Mentally Ill scale valid for use in the investigation of European nurses' attitudes towards the mentally ill? A confirmatory factor analytic approach. Journal of Advanced Nursing. 2011; 68(2): 460–70. doi: 10.1111/j.1365-2648.2011.05739.x.

139. Morrison JK, Becker BE. Seminar-induced change in a community psychiatric team's reported attitudes toward "mental illness". Journal of Community Psychology. 1975; 3(3): 281- 4.

140. Moses T. Stigma and self-concept among adolescents receiving mental health treatment. American Journal of Orthopsychiatry. 2009; 79(2): 261-74. doi: 10.1037/a0015696.

141. Nevid JS, Morrison J. Attitudes toward mental illness: the construction of the libertarian mental health ideology scale. Journal of Humanistic Psychology. 1980; 20: 71-85. doi: 10.1177/002216788002000207.

142. Ng P, Chan K. Sex differences in opinion towards mental illness of secondary school students in Hong Kong. International Journal of Social Psychiatry. 2000; 46(2): 79-88. doi: 10.1177/002076400004600201.

143. Penn DL, Guynan K, Daily T, Spaulding WD, Garbin CP, Sullivan M. Dispelling the stigma of schizophrenia: what sort of information is best? Schizophrenia Bulletin. 1994; 20(3): 567-77.

144. Pinto MD, Hickman R, Logsdon MC, Burant C. Psychometric evaluation of the revised attribution questionnaire (r-AQ) to measure mental illness stigma in adolescents. Journal of Nursing Measurement. 2012; 20(1): 47-58. doi:10.1891/1061-3749.20.1.47.

145. Ritsher JB, OtilingamaPG, Grajales M. Internalized stigma of mental illness: psychometric properties of a new measure. Psychiatry Research. 2003; 121: 31–49. doi:10.1016/j.psychres.2003.08.008.

146. Schneider J, Beeley C, Repper J. Campaign appears to influence subjective experience of stigma. Journal of Mental Health. 2011; 20(1): 89-97. doi:10.3109/09638237.2010.537403

147. Struening EL, Cohen J. Factorial invariance and other psychometric characteristics of five opinions about mental illness factors. Educational and Psychological Measurement. 1963; 23(2): 289-98. doi: 10.1177/001316446302300206.

148. Struening EL, Perlick DA, Link BG, Hellman FH, Herman D, Sirey JA. The extent to which caregivers believe most people devalue consumers and their families. Psychiatric Services. 2001; 52(12): 1633-8.

149. Stuart H, Milev R, Koller M. The inventory of stigmatizing experiences: its development and reliability. World Psychiatry. 2005; Suppl 4: 35-9.

150. Svensson B, Markström U, Bejerholm U, Björkman T, Brunt D, Eklund M, et al. Test - retest reliability of two instruments for measuring public attitudes towards persons with mental illness. BMC Psychiatry. 2011; 11:11. http://www.biomedcentral.com/1471-244X/11/11. Accessed 14 Aug 2015.

151. Swami V, Furnham A. Preliminary examination of the psychometric properties of the Psychiatric Scepticism Scale. Scandinavian Journal of Psychology. 2011; 52: 399-403. doi: 10.1111/j.1467-9450.2011.00881.x.

152. Świtaj P, Paweł Grygiel P, Jacek Wciórka J, Humenny G, Anczewska M. The stigma of subscale of the Consumer Experiences of Stigma Questionnaire (CESQ): a psychometric evaluation in Polish psychiatric patients. Comprehensive Psychiatry. 2013; 54: 713-9. doi: 10.1016/j.comppsych.2013.03.001.

153. Taylor SM, Dear MJ. Scaling community attitudes toward the mentally ill. Schizophrenia Bulletin. 1981; 7(2): 226-40.

154. Vega WA, Rodriguez MA, Ang A. Addressing stigma of depression in Latino primary care patients. General Hospital Psychiatry. 2010; 32: 182-91. doi:10.1016/j.genhosppsych.2009.10.008.

155. Vogel DL, Wade NG, Haake S. Measuring the self-stigma associated with seeking psychological help. Journal of Counseling Psychology. 2006; 53(3): 325-37.doi: 10.1037/0022-0167.53.3.325.

156. Vogel DL, Wade NG, Ascheman PL. Measuring perceptions of stigmatization by others for seeking psychological help: reliability and validity of a new stigma scale with college students. Journal of Counseling Psychology. 2009; 56(2): 301-8. doi: 10.1037/a0014903.

157. Vogt D, Di Leone BAL, Wang JM, Sayer NA, Pineles SL. Endorsed and anticipated stigma inventory (EASI): a tool for assessing beliefs about mental illness and mental health treatment among military personnel and veterans. Psychological Services. 2014; 11(1): 105-13. doi: 10.1037/a0032780.

158. Watson AC, Miller FE, Lyons JS. Adolescent attitudes toward serious mental illness. The Journal of Nervous and Mental Disease. 2005; 193(11): 769-72. doi:10.1097/01.nmd.0000185885.04349.99.

159. Wu TH, Chang CC, Chen CY, Wang JD, Lin CY. Further psychometric evaluation of the self-stigma scale – short: measurement invariance across mental illness and gender. PLoS ONE. 2015; 10(2): e0117592. doi:10.1371/journal.pone.0117592.

160. Yamaguchi S, Koike S, Watanabe K, Ando S. Development of a Japanese version of the reported and intended behavior scale: reliability and validity. Psychiatry and Clinical Neurosciences. 2014; 68: 448-55. doi: 10.1111/pcn.12151.

161. Zisman-Ilani Y, Levy-Frank-Levy I, Hasson-Ohayon I, Kravetz S, Mashiach-Eizenberg M, Roe D. Measuring the internalized Stigma of parents of persons with a serious mental illness. Journal of Nervous and Mental Disease. 2013;201(3): 183-87.

162. Fischer EH, Turner JL. Orientations to seeking professional help: development and research utility of an attitude scale. J Consult Clin Psychol.1970; 35(1): 79-90.

163. Cepeda-Benito A, Short P. Self-concealment, avoidance of psychological services, and perceived likelihood of seeking professional help. Journal of Counseling Psychology. 1998; 45(1): 58-64.

164. Deane FP, Wilson CJ, Ciarrochi J. Suicidal ideation and help-negation: not just hopelessness or prior help. J Clin Psychol. 2001; 57(7): 901-14.

165. Elhai JD, Schweinle W, Anderson SM. Reliability and validity of the attitudes toward seeking professional psychological help scale – short form. Psychiatric Research. 2008; 159: 320-9.

166. Fischer EH, Farina A. Attitudes toward seeking professional psychological help: a shortened form and considerations for research. Journal of College Student Development. 1995; 36(4): 368-73.

167. Hatchett G. Additional validation of the attitudes toward seeking professional psychological help scale. Psychological Reports. 2006; 98: 279-84.

168. Jorm AF, Blewitt KA, Griffiths KM, Kitchener BA, Parslow RA. Mental health first aid responses of the public: results from an Australian national survey. BMC Psychiatry. 2005; 5: 9. doi: 10.1186/1471-244X-5-9.

169. Jorm AF, Mackinnon A, Christensen H, Griffiths KM. Structure of beliefs about the helpfulness of interventions for depression and schizophrenia. Soc Psychiatry Psychiatr Epidemiol. 2005; 40: 877-83.

170. Lee J, Friesen BJ, Walker JS, Colman D, Donlan WE. Youth's help-seeking intentions for ADHD and depression: findings from a national survey. J Child Fam Stud. 2014; 23: 144-56. doi: 10.1007/s10826-012-9700-3.

171. Mackenzie CS, Knox VJ, Gekoski WL, Macaulay HL. An adaptation and extension of the attitudes toward seeking professional psychological help scale. Journal of Applied Social Psychology. 2004; 34(11): 2410-35.

172. Nickerson KJ, Helms JE, Terrell F. Cultural mistrust, opinions about mental illness, and black students' attitudes toward seeking psychological help from white counselors. Journal of Counseling Psychology.1994; 41(3): 378-85.

173. Reavley NJ, Morgan AJ, Jorm AF. Development of scales to assess mental health literacy relating to recognition of and interventions for depression, anxiety disorders and schizophrenia/psychosis. Australian & New Zealand Journal of Psychiatry. 2014; 48(1): 61-9. doi: 10.1177/0004867413491157.

174. Royal KD, Thompson JM. Measuring protestant Christians' willingness to seek professional psychological help for mental illness: a Rasch measurement analysis. Journal of Psychology and Christianity. 2012; 31(3): 195-204.

175. Sahin H, Uyar SY. Turkish high school students' attitudes towards seeking professional psychological help. US-China Education Review B 2. 2011; 251-9.

176. Schmeelk-Cone K, Pisani AR, Petrova M, Wyman PA. Three scales assessing high school students' attitudes and perceived norms about seeking adult help for distress and suicide concerns. Suicide and Life-Threatening Behavior. 2012; 42(2): 157-72. doi: 10.1111/j.1943-278X.2011.00079.x.

177. Turner E. The parental attitudes toward psychological services inventory: adaptation and development of an attitude scale. Community Mental Health Journal. 2012; 48: 436-49. doi: 10.1007/s10597-011-9432-7.

178. Wilson CJ, Deane FP, Rickwood D. Measuring help-seeking intentions: properties of the general help-seeking questionnaire. Canadian Journal of Counseling. 2005; 39(1): 15-28.

179. Pinfold V, Toulmin H, Thornicroft G, Huxley P, Farmer P. Reducing psychiatric stigma and discrimination: evaluation of educational interventions in UK secondary schools. British Journal of Psychiatry. 2003; 182: 342-6. doi: 10.1192/bjp.02.375.

180. Smith JV, Birchwood MJ. Specific and non-specific effects of educational intervention with families living with a schizophrenic relative. British Journal of Psychiatry.1987; 150: 645-52. doi: 10.1192/bjp.150.5.645.

181. Wood A, Wahl O. Evaluating the effectiveness of a consumer-provided mental health recovery education presentation. Psychiatric Rehabilitation Journal. 2006; 30(1): 46-52.

182. Stuart H, Arboleda-Florez J. Community attitudes toward people with schizophrenia. Canadian Journal of Psychiatry. 2001; 46: 245-52.

183. Ghanizadeh A, Bahredar MJ, Moeini SR. Knowledge and attitudes towards attention deficit hyperactivity disorder among elementary school teachers. Patient Education and Counseling. 2006; 63: 84-8. doi:10.1016/j.pec.2005.09.002.

184. DeSocio J, Stember L, Schrinsky J. Teaching children about mental health and illness: A school nurse health education program. The Journal of School Nursing. 2006; 22(2): 81-6.

185. Essler V, Arthur A, Stickley T. Using a school-based intervention to challenge stigmatizing attitudes and promote mental health in teenagers. Journal of Mental Health. 2006; 15(2): 243-50. doi: 10.1080/09638230600608669.

186. Hoven C, Doan T, Musa GJ, Jaliashvili T, Duarte CS, Ovuga, et al. Worldwide child and adolescent mental health begins with awareness: A preliminary assessment in nine countries. International Review of Psychiatry. 2008; 20(3): 261–70.doi: 10.1080/09540260801995950.

187. Ojanen M. Attitudes towards Mental Patients. International Journal of Social Psychiatry.1992; 38(2): 120-30. doi: 10.1177/002076409203800205.

188. Mullen A, Murray L, Happell B. Multiple family group interventions in first episode psychosis: Enhancing knowledge and understanding. International Journal of Mental Health Nursing. 2002; 11: 225-32.

189. Williams B, Pow J. Gender differences and mental health: an exploratory study of knowledge and attitudes to mental health among Scottish teenagers. Child and Adolescent Mental Health. 2007; 12(1): 8–12. doi: 10.1111/j.1475-3588.2006.00413.x.

190. Cabassa LJ, Lester R, Zayas LH. "It's Like Being in a Labyrinth:" Hispanic immigrants' perceptions of depression and attitudes toward treatments. J Immigr Minor Health. 2007; 9(1): 1–16. doi:10.1007/s10903-006-9010-1.

191. Link BG, Phelan JC, Bresnahan M, Stueve A, Pescosolido BA. Public conceptions of mental illness: labels, causes, dangerousness, and social distance. American Journal of Public Health.1999; 89(9): 1328-33.

192. Kelly AE, Achter JA. Self-concealment and attitudes toward counseling in university students. Journal of Counseling Psychology. 1995; 42(1): 40-6.

193. Hugo CJ, Boshoff DEL, Traut A, Zungu-Dirwayi N, Stein DJ. Community attitudes toward and knowledge of mental illness in South Africa. Soc Psychiatry Psychiatr Epidemiol. 2003; 38: 715–9. doi: 10.1007/s00127-003-0695-3.

194. Ozmen E, Ogel K, Aker T, Afsın Sagduyu A, Tamar D, Boratav C. Public opinions and beliefs about the treatment of depression in urban Turkey. Soc Psychiatry Psychiatr Epidemiol. 2005; 40: 869–76. doi: 10.1007/s00127-005-0985-x.

195. Iloabachie C, Wells C, Goodwina B, Baldwin M, Vanderplough-Booth K, Gladstone T, et al. Adolescent and parent experiences with a primary care/Internet-based depression prevention intervention (CATCH-IT). General Hospital Psychiatry. 2011; 33: 543–55. doi:10.1016/j.genhosppsych.2011.08.004.

196. Costin DL, Mackinnon AJ, Griffiths KM, Batterham PJ, Bennett AJ, Bennett K, Christensen H. Health e-Cards as a means of encouraging help seeking for depression among young adults: randomized controlled trial. J Med Internet Res. 2009;11(4): e42. doi:10.2196/jmir.1294.

197. Nguyen QCX, Anderson LP.Vietnamese Americans' attitudes toward seeking mental health services: relation to cultural variables. Journal of Community Psychology. 2005; 33(2): 213-31. doi: 10.1002/jcop.20039.

198. Teng EJ, Friedman LC. Increasing mental health awareness and appropriate service use in older Chinese Americans: A pilot intervention. Patient Education and Counseling. 2009; 76: 143–6. doi:10.1016/j.pec.2008.11.008.

199. Santor DA, Poulin C, LeBlanc JC, Kusumakar V. Facilitating help seeking behavior and referrals for mental health difficulties in school aged boys and girls: a school-based intervention. Journal of Youth Adolescence. 2007; 36: 741–52. doi: 10.1007/s10964-006-9092-z.

200. Brown C, O'Conner K, Copeland VC, Grote N, Beach S, Battista D, & Reynolds CF. Depression stigma, race, and treatment seeking behavior and attitudes. Journal of Community Psychology. 2010; 38(3), 350-358. doi: 10.1002/jcop.20368

201. Askell-Williams H, Lawson M, Murray-Harvey R. Teaching and learning about mental illnesses. International Journal of Mental Health Promotion. 2007; 9(4): 26-36.

202. Chang H. Help-seeking for stressful events among Chinese college students in Taiwan: roles of gender, prior history of counseling, and help-seeking attitudes. Journal of College Student Development. 2008; 49(1): 41-51. doi: 10.1353/csd.2008.0003.

203. Sussman LK, Robins LN, Earls F. Treatment-seeking for depression by black and white Americans. Soc. Sci. Med. 1987; 24(3): 187-96.

204. Pajares F, Urdan T. The ones we remember: scholars reflect on teachers who made a difference. In: Kirshner B, editor. Adolescence and education. Boulder: University of Colorado School of Education; 2008.

205. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60: 34-42.

206. Pescosolido B, Jensen P, Martin J, Perry B, Olafsdottir S, Fettes D. Public knowledge and assessment of child mental health problems: Finding form the national stigma study – children. Journal of the American Academy of Child and Adolescent Psychiatry. 2008; 47(3): 339-49.

207. Thornicroft G. Shunned: discrimination against people with mental illness. London: OUP Oxford; 2006.

208. Jones EE, Farina A, Hastorf AH, Marcus H, Miller DT, Scott RA. Social stigma: the psychology of marked relationships. New York, NY: Freeman and Company; 1984.

209. Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall; 1980.

210. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O et al. Child and adolescent mental health worldwide evidence for action. Lancet. 2011; 378: 1515-1525.

211. Waddell C, Offord DR, Sheperd CA, Hua JM, & McEwan K. Child psychiatric epidemiology and Canadian public policy-making: the state of the science and the art of the possible. Canadian Journal of Psychiatry. 2002; 47(9): 852-832.

212. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. Am Psychol. 2012; 67(3):231–243.

213. Wei Y, Kutcher S, Szumilas M. Comprehensive school mental health: an integrated "school-based pathway to care" model for Canadian secondary schools. McGill J Educ. 2011; 46(2): 213–229.

214. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012; 21(4):651-7.

215. Uijen AA, Heinst CW, Schellevis FG, van den *Bosch* WJ, van de Laar FA, Terwee CB, & Schers, HJ. Measurement properties of questionnaires measuring continuity of care: a systematic review. PloS ONE. 2012; 7 (7): e42256.

216. Furlan AD, Pennick V, Bombardier C, van TM. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976). 2009; 34 (18): 1929-1941. 21

217. Balasubramanian N, Juliana LD, Sathyanarayana RTS. Knowledge questionnaire on home care of schizophrenics (KQHS): validity and reliability. Journal of Education and Practice. 2013; 4(11): 176-182.

218. Daltio CS, Attux C, Ferraz MB. Knowledge in schizophrenia: The Portuguese version of KAST (Knowledge About Schizophrenia Test) and analysis of social-demographic and clinical factors' influence. Schizophrenia Research. 2015; 168:168-173.

219. O'Connor M & Casey L. The Mental Health Literacy Scale (MHLS): a new scale-based measure of mental health literacy. Psychiatry Research. 2015; 229: 511-516.

220. Hart SR, Kastelic EA, Wilcox HC, Beaudry MB, Musei RJ, Heley K et al. Achieving depression literacy: the adolescent depression knowledge questionnaire (ADKQ). Sch Ment Health. 2014; 6:213–23. doi:10.1007/s12310-014-9120-1

221. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003; 12:349-62.

222. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; 16(3): 297-334.

223. Streiner DL, & Norman GR. Health measurement scales: a practical guide to their development and use, fourth edition. Oxford University Press, New York; 2008.

224. Centre for Chronic Disease Prevention, Public Health Agency of Canada. Positive Mental Health Surveillance Indicator Framework, 2016 Edition. 2016. Available from: http://infobase.phac-aspc.gc.ca/positive-mental-health/. [Accessed July 28, 2016]

225. Breslau J, Lane M, Sampson N, Kessler RC. Mental disorders and subsequent educational attainment in a US national sample. J Psychiat Res. 2008; 42, 708-716.

226. Vigod SN, Dennis CL, Kurdyak PA, Cairney J, Guttmann A, Taylor VH. Fertility rate trends among adolescent girls with major mental illness: a population-based study. Pediatrics. 2014; 133(3): e585-91. doi: 10.1542/peds.2013-1761.

227. Milin R, Kutcher S, Lewis S, Walker S, Wei Y, Ferrill N, & Armstrong M. Impact of a mental health curriculum on knowledge and stigma among high school students: a randomized controlled trial. Journal of the American Academy of Child & Adolescent Psychiatry. 2016; 55(5):383-391. DOI: 10.1016/j.jaac.2016.02.01

228. McLuckie A, Kutcher S, Wei Y, Weaver C. Sustained improvements in students' mental health literacy with use of a mental health curriculum in Canadian schools. 2014; BMC Psychiatry 14(1), 1694. DOI: 10.1186/s12888-014-0379-4

229. Kutcher S, Wei Y, Morgan C (2015). Successful application of a Canadian mental health curriculum resource by usual classroom teachers in significantly and sustainably improving student mental health literacy. 2015; Canadian Journal of Psychiatry 60(12), 580-586.

230. Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J of Clin Epidemiol. 2009; 62: 1006-1012.

231. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, & de Vet HCW. COSMIN checklist manual. Available from: http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf. [Accessed 11 June, 2016].

232. Assefa D, Shibre T, Asher L, & Fekadu A. Internalized stigma among patients with schizophrenia in Ethiopia: a cross-sectional facility-based study. BMC Psychiatry. 2012; 12: 239. http://www.biomedcentral.com/1471-244X/12/239

233. Baker JA, Richards DA, & Campbell M. Nursing attitudes towards acute mental health care: development of a measurement tool. J Adv Nurs. 2005; 49(5): 522-529.

234. Brockington IF, Hall P, Levings J, & Murphy C. The community's tolerance of the mentally ill. Bri J of Psychiat. 1993; 162: 93-99.

235. Brohan E, Gauci D, Sartorius N, Thornicroft G, & For the GAMIAN-Europe Study Group. Self-stigma, empowerment and perceived discrimination among people with bipolar disorder or depression in 13 European countries: the GAMIAN-Europe study. J Affect Disorders. 2011; 129: 56-63.

236. Corrigan PW, Watson AC, Warpinski AC, & Gracia G. Stigmatizing attitudes about mental illness and allocation of resources to mental health services. Community Ment HLT J. 2004; 40(4): 297-307.

237. Haddad M, Walters P, Tylee A. District nursing staff and depression: a psychometric evaluation of depression attitude questionnaire findings. Int J Nurs stud. 2007; 44: 447-456.

238. Link BG, Cullen FT, Struening E, Shrout PE, Dohrenwend BP. A modified labeling theory approach to mental disorders: an empirical assessment. Am Sociol Rev. 1989; 54(3): 400-423.

239. Link BG, Mirotznik J, & Cullen FT. The effectiveness of stigma coping orientations: can negative consequences of mental illness labeling be avoided?" J Health and Soc Behav. 1991; 32:302-20.

240. Link BG, Struening EL, Rahav M, Phelan JC, Nuttbrock L. On stigma and its consequences: evidence from a longitudinal study of men with dual diagnoses of mental illness and substance abuse. J Health Soc Behav. 1997; 38(2): 177-190.

241. Madianos MG, Madianou D, Vlachonikolis J, & Stefanis CN. Attitudes towards mental illness in the Athens area: implications for community mental health intervention. Acta Psychiatr Scand. 1987; 75: 158-165.

242. Mak WWS, & Cheung RYM. Self-stigma among concealable minorities in Hong Kong: conceptualization and unified measurement. Am J Orthopsychiat. 2010; 80(2): 267-281. doi: 10.1111/j.1939-0025.2010.01030.x

243. Ritsher JB, & Phelan JC. Internalized stigma predicts erosion of morale among psychiatric outpatients. Psychiatry Res. 2004; 129: 257-265.

244. Sevigny R, Yang W, Zhang P, Marleau JD, Yang Z, Lin S, Li G, Xu D, Wang Y, & Wang H. Attitudes toward the mentally ill in a sample of professionals working in a psychiatric hosptital in Beijing. Int J Soc Psychiatr. 1999; 45: 41. doi: 10.1177/002076409904500106

245. Scheerder G, De Coster I, Van Audenhove, C. Community pharmacists' attitude toward depression: a pilot study. Res Soc and Adm Pharm. 2009; 5: 242-252.

246. Sibitz I, Amering M, Unger A, Seyringer ME, Bachmann A, Schrank B, Benesch T, Schulze B, Woppmann A. The impact of the social network, stigma and empowerment on the quality of life in patients with schizophrenia. Eur Psychiat. 2011; 26: 28-33.

247. Sibitz I, Unger A, Woppmann A, Zidek T, & Amering M. Stigma resistance in patients with schizophrenia. Schizophrenia Bull. 2011; 37(2): 316-323.

248. Sorsdahl KR, Kakuma R, Wilson Z, Stein DJ. The internalized stigma experienced by members of a mental health advocacy group in South Africa. Int J Soc Psychiat. 2012; 58: 55. doi: 10.1177/0020764010387058.

249. Thornicroft G, Brohan E, Rose D, Sartorius N, Leese M, for the INDIGO Study Group. Global pattern of experienced and anticipated discrimination against people with schizophrenia: a cross-sectional survey. Lancet. 2009; 373: 408-15.

250. Wolff G, Pathare S, Graig T, & Leff J. Community attitudes to mental illness. Brit J Psychiat. 1996; 168: 183-190.

251. Maccoby, Eleanor E. and Nathan Maccoby. "The Interview: A Tool of Social Science." In: Lindzey G (ed) Handhook of Social Psychology. Vol. I. Cambridge, MA: Addison-Wesley, pp 449-487; 1954.

252. Canadian Agency for Drugs and Technologies in Health Information Services. Grey Matters: a practical tool for searching health-related grey literature. 2015. Available from: https://www.cadth.ca/resources/finding-evidence/grey-matters. [Accessed 07 August, 2016].

253. Ang RP, Lau S, Tan AG, & Lim KM.Refining the attitudes toward seeking professional psychological help scale: factorial invariance across two Asian samples. Measurement and Evaluation in Counseling and Development. 2007;40, 130-141.

254. Cakar FS & Savi S. An exploratory study of adolescent's help-seeking sources. Social and Behavioral Sciences. 2014;159, 610 – 614.

255. Han DY & Chan SH. Is the Attitudes Toward Seeking Professional Psychological Help scale applicable to ethnic Chinese students? Psychometric properties and cultural considerations. International Research. 2015; 56(1), 73-83.

256. Tuliao AP. &Velasquez PA. Revisiting the general help seeking questionnaire: adaptation, exploratory factor analysis, and further validation in a Filipino college student Sample. Philippine Journal of Psychology. 2014; 47(1), 1-17.

257. D'Anstiss H, Ziaian T, Procter N, Warland J, &Baghurst P. Help-seeking for mental health problems in young refugees: a review of the literature with implications for policy, practice, and research. Transcultural Psychiatry. 2009; 46(4), 584-607. doi: 10.1177/1363461509351363

258. Susukida R, Mojtabai R, &Mendelson T. Sex Differences in help seeking for mood and anxiety disorders in the national comorbidity survey-replication. Depression and Anxiety. 2015; 32(11), 853-60. doi: 10.1002/da.22366. Epub. 2015 Apr 22.

259. Ajzen I. From intentions to actions: a theory of planned behavior. In J. Kuhl & J. Beckman (Eds). Action-Control: from Cognition to Behavior (pp. 11-39). Heidelberg, Germany: Springer; 1985.

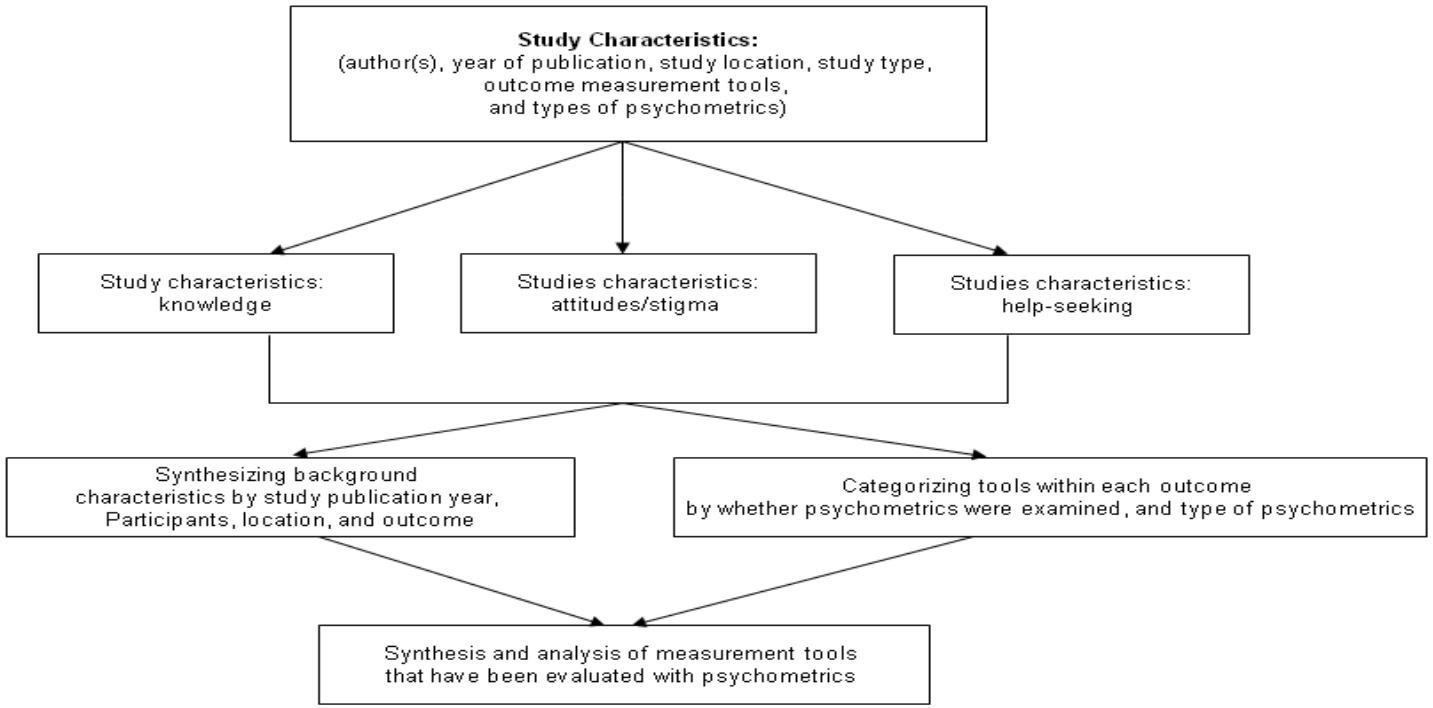
260. Livingston JD, Tugwell A, Korf-Uzan K, Cianfrone M, & Coniglio C. Evaluation of a campaign to improve awareness and attitudes of young people towards mental health issues. Social Psychiatry & Psychiatric Epidemiology. 2013; 48, 965–973. doi 10.1007/s00127-012-0617-3

APPENDICES

Appendix 1: Figures and Tables

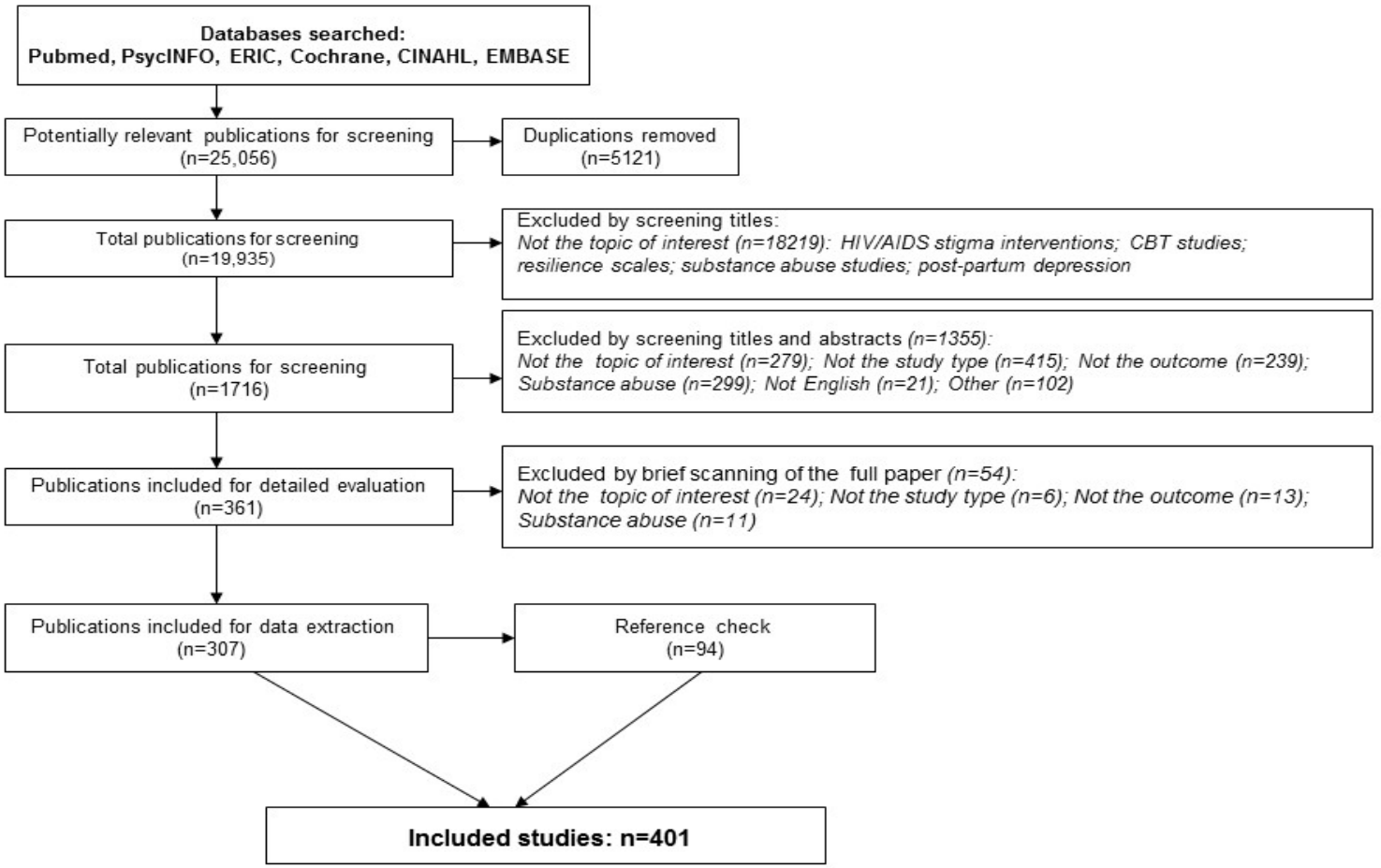
Chapter 2 Figure 1: Charting Process (Data Extraction Process)

Figure 1: Charting process (data extraction process)

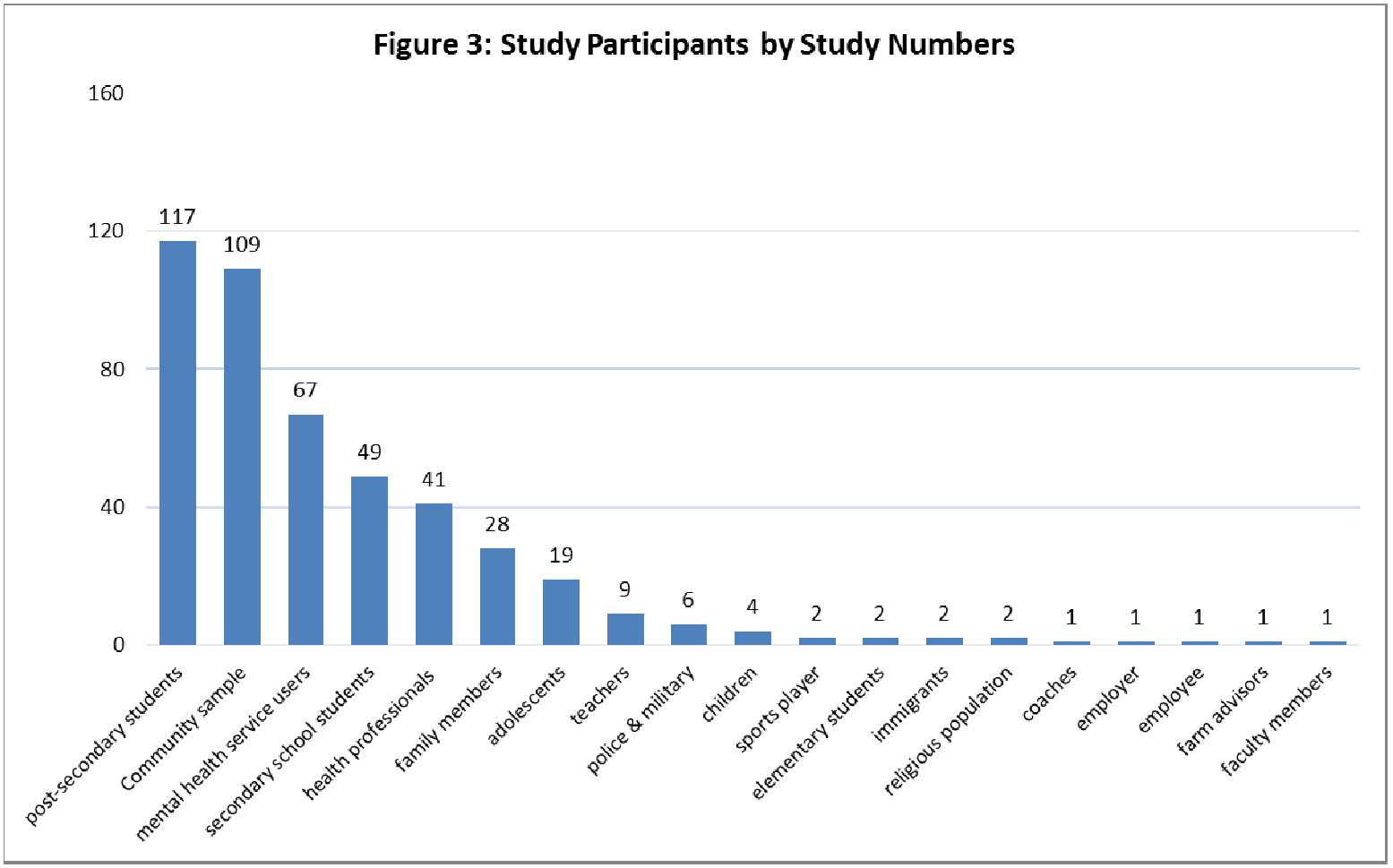


Chapter 2 Figure 2: Search Results

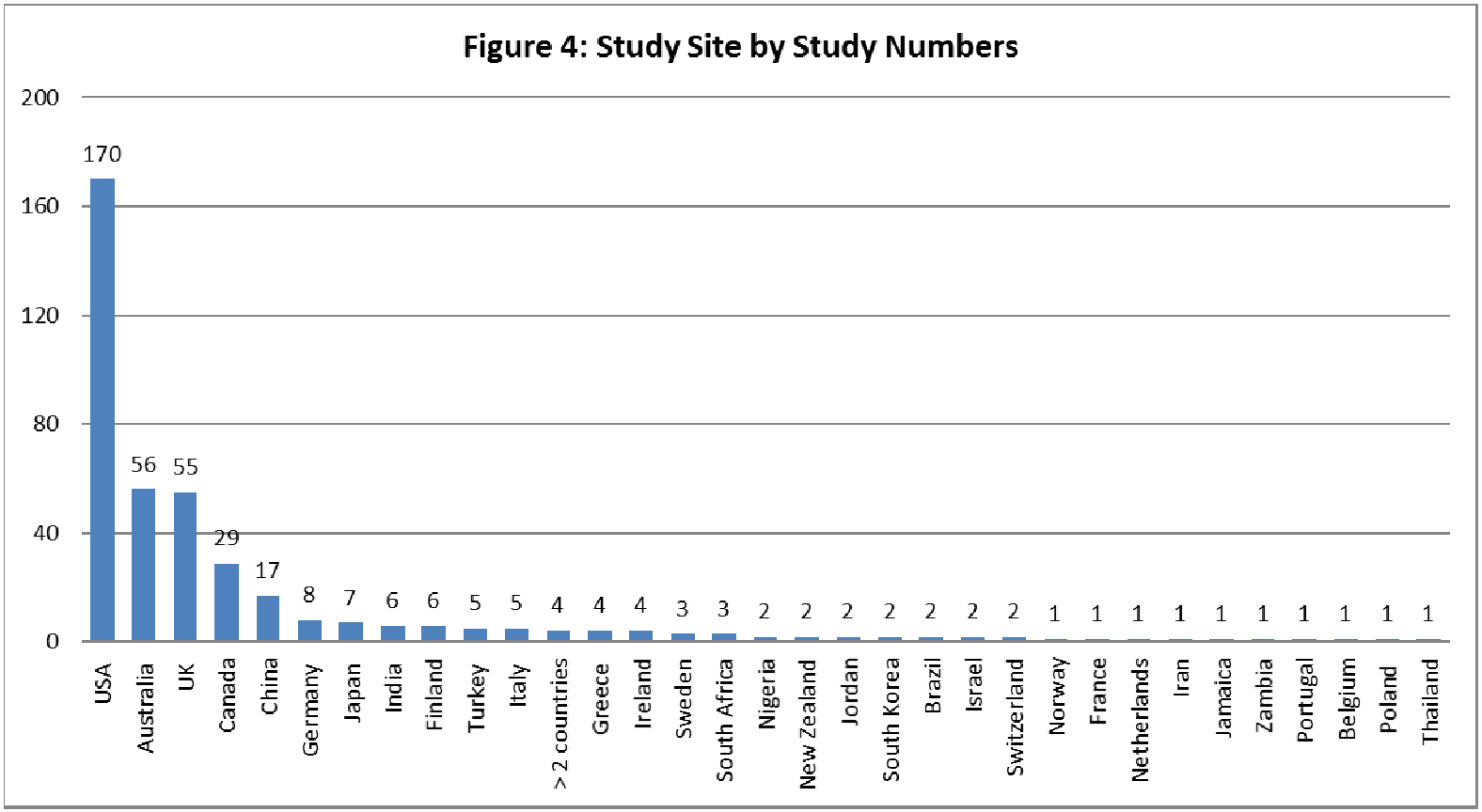
Figure 2: Search results



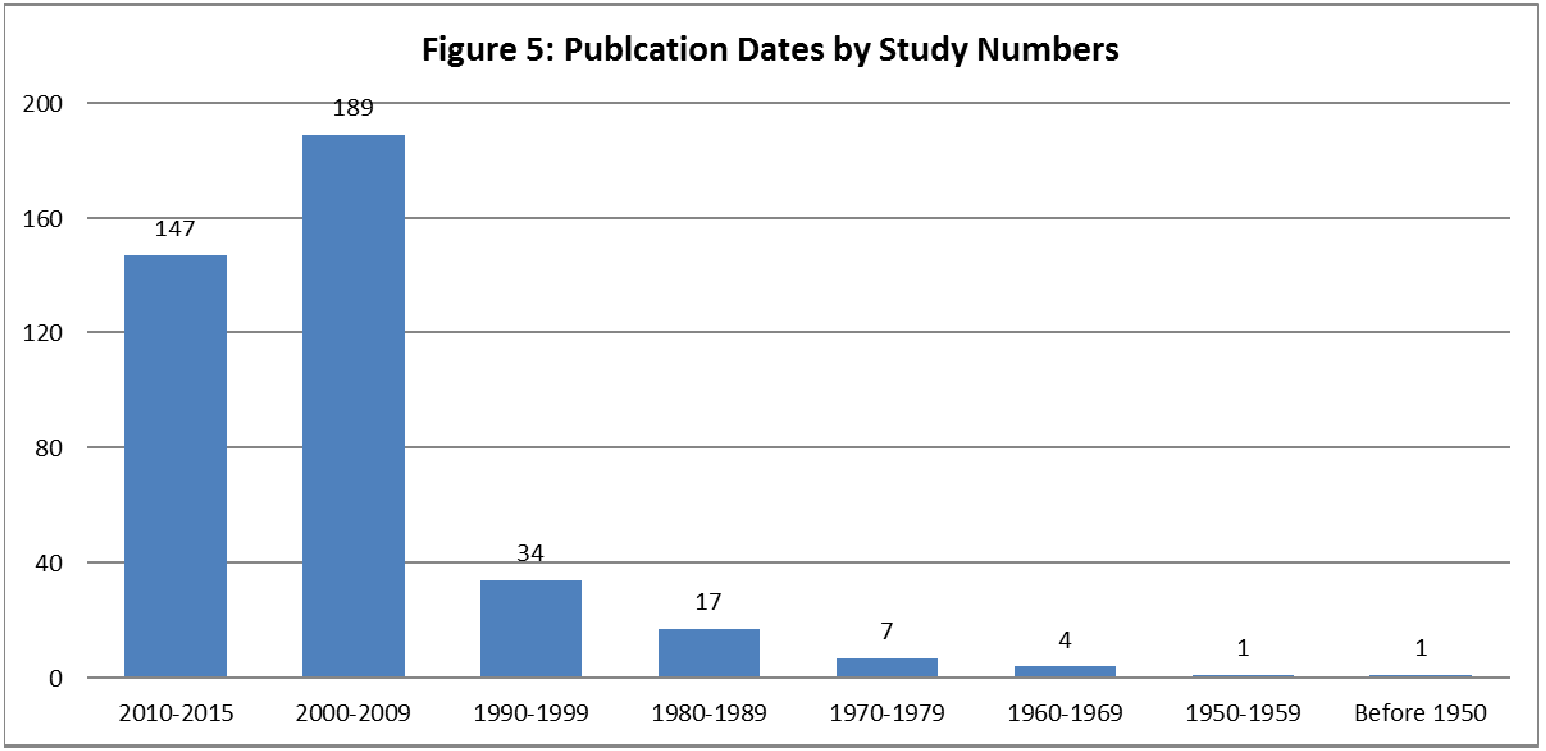
Chapter 2 Figure 3: Study Participants by Study Numbers



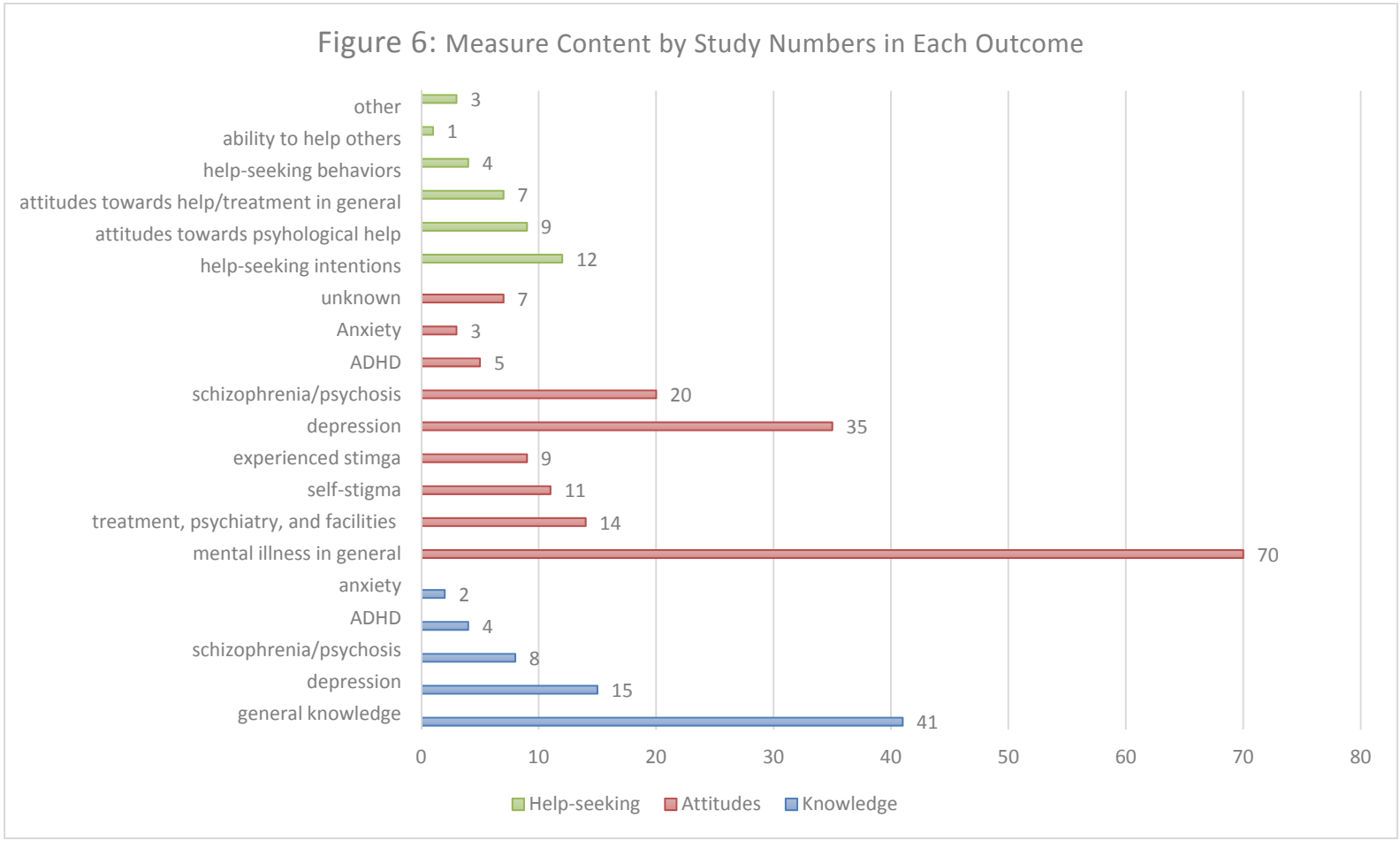
Chapter 2 Figure 4: Study Site by Study Numbers



Chapter 2 Figure 5: Publication Dates by Study Numbers

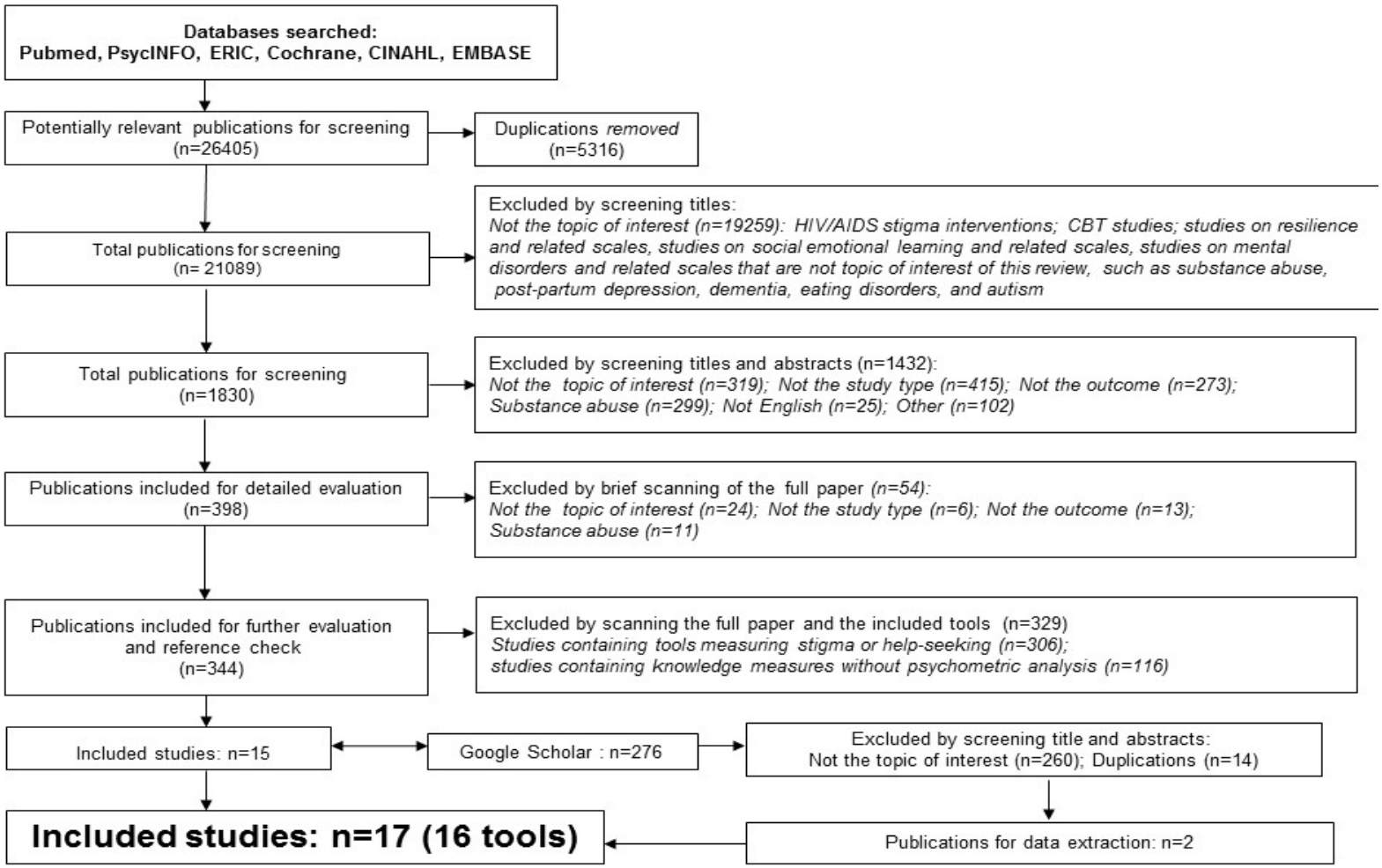


Chapter 2 Figure 6: Measure Content by Study Numbers in Each Outcome

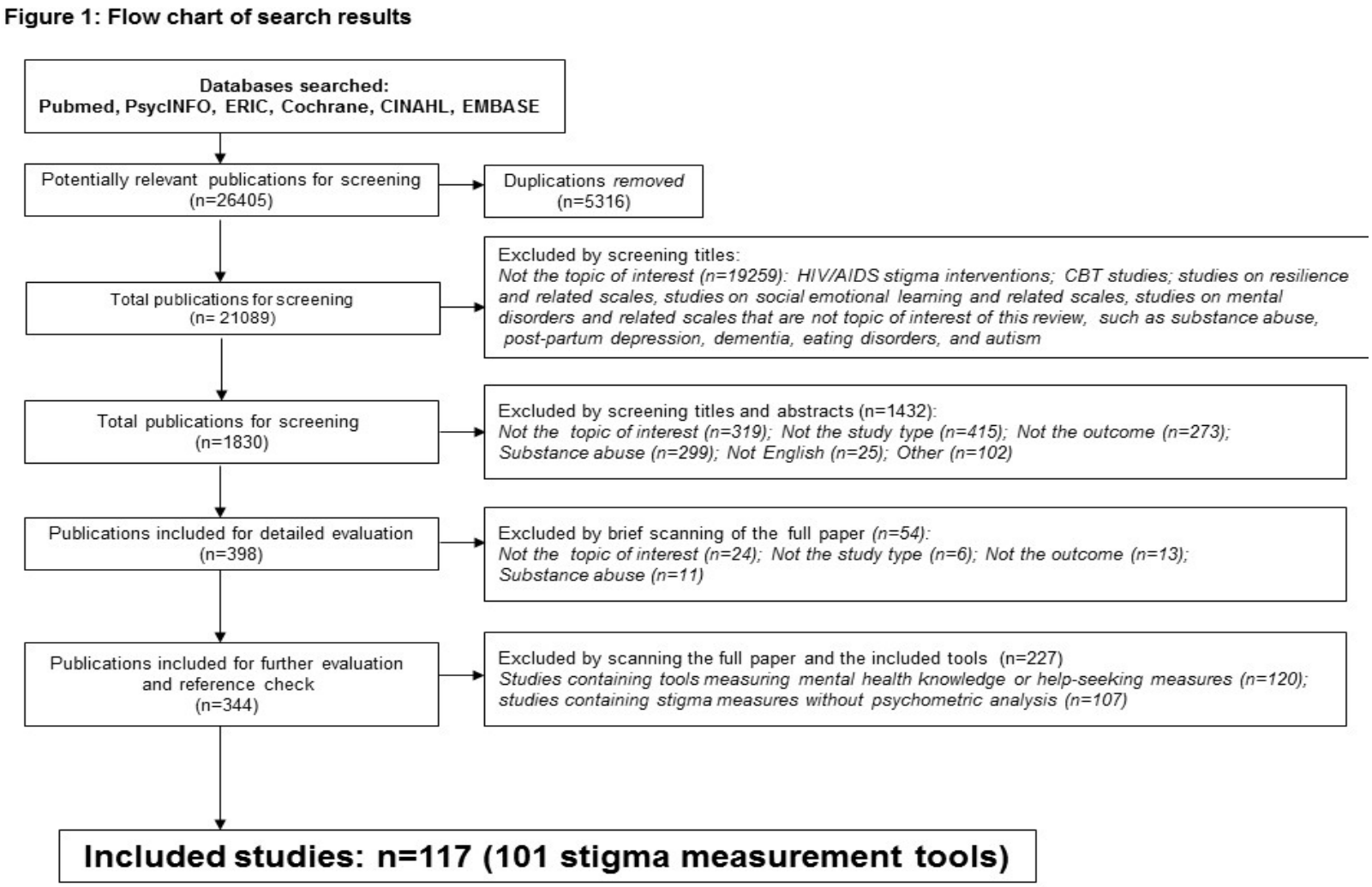


Chapter 3 Figure 1: Flow Chart of Search Results

Figure 1: Flow chart of search results

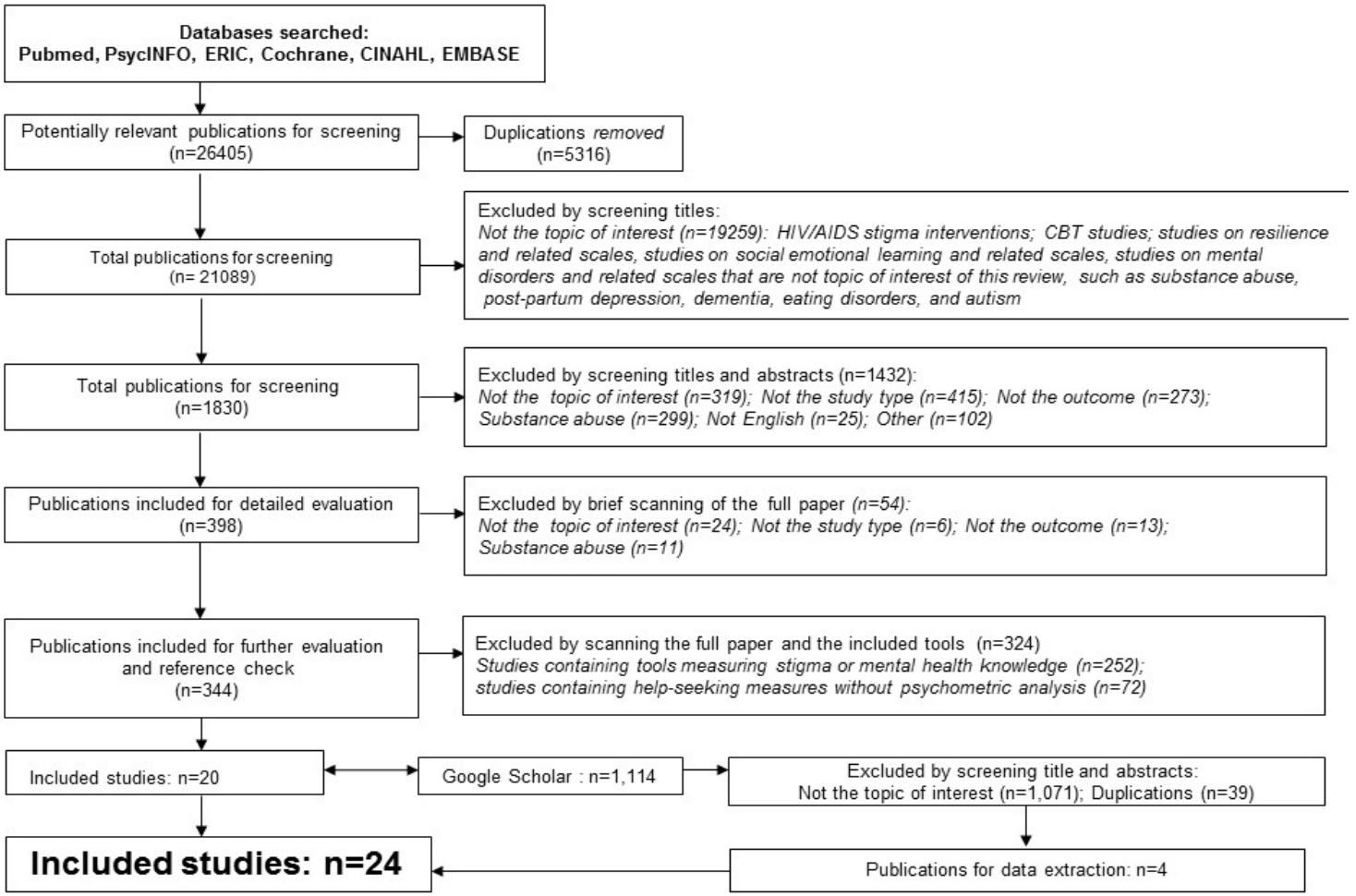


Chapter 4 Figure 1: Flow Chart of Search Results



Chapter 5 Figure 1: Flow Chart of Search Results

Figure 1: Flow chart of search results



TABLES

Chapter 2 Table 1: Psychometrics of Knowledge Measures

Measures	Developer/Author Reliability		Validity & Responsiveness	Content
1. Knowledge about Schizophrenia Questionnaire (KASQ)	Ascher-Svanum, 1999 [48]	KR-20*= .85, .89 r it*=.46; .51 r**= .83 (p<.005)	R; CT	S
2. Knowledge About Schizophrenia Test (KAST)	Compton et al., 2007 [49]	KR-20*=.82 (.45-.78)	CS; CT; CR	S
3. Multiple-Choice Knowledge of Mental Illnesses Test (MC-KOMIT)	Compton et al., 2011 [50]	α*=.68-.75 r**=.79 (p<.001)	R; CS; CT	G
4. Mental Health Knowledge Schedule (MAKS);	Evans-Lacko et al., 2010 [51]	α*=.65 Lin's Pc **=.71; k**=.57-.87	CT	G
5. Depression Multiple Choice Question (MCQ)	Gabriel & Violato, 2009 [52]	α*=.68	CT; CV; FA	D
6. Depression literacy scale (D-Lit)	Kiropoulos et al., 2011 [53]	α*=..88, .92 r**=.78, .80 (p<0.001)		D
	Gulliver et al., 2012 [54]	α*=.70 r**= .71 (p = .02)		
7. Anxiety Literacy Questionnaire (A-Lit)	Gulliver et al., 2012 [54]	α*=.76; r**= .83 (p=.003)		A
8. Test of Knowledge About ADHD (KADD)	Hepperlen et al., 2002 [55]	α*=.81-.82;	FA; CT	ADHD
9. Knowledge about Depression and Mania Inventory (KDMI)	Kronmuller et al., 2008 [56]	α*=.89 (.76-.81); r it* = .36-.43	CC/CR; CT; D; R	D
10. Journey of Hope (JOH) Outcome Survey	Pickett-Schenk et al., 2000 [57]	α*=.75-.83	CS; FA	G

Measures	Developer/Author Reliability		Validity & Responsiveness	Content
11. Knowledge of Mental Disorders (KMD)	Serra et al., 2013 [58]	α*=.588	CS; FA	G
12. Adolescent Depression Knowledge Questionnaire (ADKQ)	Shelley et al., 2014 [59]	α*=.89	FA	D
13. Mental health disorder recognition questionnaire (MDRQ)	Swami et al., 2011 [60]	k***=.94, .96	CS; CV	G
14. Mental Health Knowledge Questionnaire (MHKQ)	Wang et al., 2013 [61]	α*=.69	FA	G

1. * internal consistency reliability: Cronbach alpha (α), Kuder-Richardson 20 (KR-20), item to total correlation (r it) ; **test-retest reliability: weighted kappa (k), Pearson correlation coefficient (r), Lin's Pc; *** inter-rater reliability: weighted kappa (k)
2. Content validity: CT; Construct validity: CS; Criterion validity: CR; Convergent validity: CV; Concurrent validity: CC; Discriminant validity: D; Factor analysis: FA; Responsiveness: R.
3. G: general knowledge; D: Depression; S: Schizophrenia; Attention Deficit Hyperactivity Disorder: ADHD; A: Anxiety

Chapter 2 Table 2: Psychometrics of Stigma/Attitudes Measures

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
1. Social Distance (SD)	Bogardus, 1925 [68]			
Link Social Distance scale (1987)	Link et al., 1987 [125]	α*=.74; .75; .92	CS; CR/CV; D; FA	A
Bogardus Social Distance Scale (modified)	Angermeyer & Matschinger, 2003 [63]	α*=.90	FA	
Social distance scale	Link, 1983 [123]	α*=.85, .91;	CS	
Reported and Intended Behaviour Scale (RIBS)	Evans-Lacko et al., 2011 [88]	α*=.85; k**=.75		
RIBS- Japanese (RIBS-J)	Yamaguchi et al., 2014 [160]	α*=.83 Lin's Pc** = .71	CC; FA	
Social Contact Scale (SCS)	Jackson & Heatherington, 2006 [110]	α=.55-.75	FA	
The Social Supports Acceptance Scale (SSAS)	Mansouri & Dowell, 1989 [133]	α=.80-.94	CS; R	
2. Opinions about Mental Illness (OMI)	Cohen & Struening, 1962 [78]		CS; FA	A
OMI	Struening & Cohen, 1963 [147]	α*=.299-.80	FA	
OMI in Chinese Community scale (OMICC)	Ng & Chan, 2000 [142]	α*=.87 (.43-.72)	FA	
3. Community Attitudes towards Mental Illness (CAMI)	Taylor & Dear, 1981 [153]	α*=.62-.90	CS; D; FA	A
Fear and Behavioural Intentions (FABI)	Svensson et al., 2011 [150]	α*=.80 k**=.29-.54		
Mental Health Attitude Survey for Police	Clayfield et al., 2011 [77]	α*=.87	CS; CV; FA	
4. Devaluation-Discrimination tool (DD)	Link 1987 [124]	α*=.73-.83	CS	A

	Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
	Perceived Discrimination Devaluation (PDD)	Interian et al., 2010 [108]	α*=.80	CS; CV/CR; FA	A
	Public stigma (PS)	Moses, 2009 [140]	α*=.76 K***=.79-.90	CV; CS; D	
	Stigma-Devaluation Scale (SDS)	Dalky, 2012 [83]	α*=.87	FA	
	Depression is a Matter of Will (DMW)	Aromaa et al., 2010 [64]		CS; FA	
5.	Depression Stigma scale (DSS)	Griffiths et al., 2004 [98]	α*=.75-.82 r**= .86 (p=.001)	DS; CV; D; FA	A
6.	Attribution Questionnaire (AQ)	Corrigan et al., 2003 [80]	α*=.70-.96	CS; FA	A
	AQ-27	Brown, 2008 [72]	α*=.60-.93 ICC**=.72-.90	CV; FA	
	r-AQ	Pinto et al., 2012 [144]	α*=.70	FA	
7.	Internalized Stigma of Mental Illness (ISMI)	Ritsher et al., 2003 [145]	α*=.84-.98 r**=.92 (.61-.91) (p<.05) ICC**=.78	CS; CC; D; FA; P	C
	Parents' Internalized Stigma of Mental Illness Scale (PISMI)	Zisman-Ilani et al., 2013 [161]	α*=.61-.78	FA	
	ISMI Chinese (ISMIS-C)	Lien et al., 2015 [122]	α*=.90; ICC=.36-.73	CS; FA	
	ISMI-10	Boyd et al., 2014 [35]	α*=.75	CT; CC; CS	
8.	Perceived dangerousness (PD)	Link, et al., 1987 [125]	α*=.85	CS	A
	Link Stigma Scale (dangerousness)	Bagley & King, 2005 [65]	α* >.80	CS; CR; D	
	Dangerousness Scale (DS)	Penn et al., 1994 [143]	α*=.78	CS	
9.	British Omnibus National Survey (ONS)	Kobau et al., 2010 [119]	α*=.66-.69	CV; CC; FA	A
	Changing Mind	Svensson et al., 2011 [150]	α*=.19-.46 r**:Poor to moderate		
10.	Self-stigma of Seeking Psychological Help (SSOSH)	Vogel et al., 2006 [155]	α*=.88 ; r**=.72	CS; P; D; FA	C

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
11. Self-stigma of Mental Illness (SSMI)	Corrigan et al., 2006 [82]	α*=.64-.91; r**=.62-.82	CS; D	C
SSMI-Short Form	Corrigan et al., 2012 [81]	α*=.22-.87	CS; D	
12. Attitudes to Mental Illness Questionnaire (AMID)	Luty et al., 2006 [127]	r**=.70; .93	CC; FA	A
13. Stigma Scale for Receiving Psychological Help (SSRPH)	Komiya et al., 2000 [120]	α*=.72	CS; CR; FA	D
14. Affective Reaction Scale (ARS)	Penn et al., 1994 [143]	α*=.86	CS	A
15. Discrimination and Stigma Scale (DISC)	Brohan et al., 2013 [71]	α*=.78; Lin's Pc **=.88, 89 (p<.001) k**=0.45-0.89 K***= .62-.97	CV; DV	B
Questionnaire on Anticipated Discrimination (QUAD)	Gabbidon et al., 2013 [91]	α*=.86; Lin's Pc*** = .81 k**=.41-.80	CV	B
16. Mental Illness: Clinician's Attitudes (MICA)	Kassam et al., 2010 [113]	α* = .79 Lin's Pc **=.80 (p<.001)	CV; DV; FA; R	A
MICA-v4	Gabbidon,J., 2013 [92]	α*=.72; r it *≥.2	CV; FA	
17. Day's Mental Illness Scale (DMIS)	Day et al., 2007 [84]	α*=.71-.86	CS; FA	A
18. ADHD Stigma Questionnaire (ASQ)	Kellison et al., 2010 [117]	α*=.55-.93; ICC**=.71(.55-.73)	CS; CV; DV; FA	A
Stigmatization towards Adults ADHD	Fuermaier et al., 2012 [90]	α*=.91 (.61-.87)	CS	
19. Rejection Experiences (RE)	Link, 1987 [124	α*=.73-.85 K***=.79-.90	CS; CR; CV; D	B

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
20. Generalized Anxiety Stigma Scale (GASS)	Griffiths et al., 2011 [100]	α*=.86-0.91 r**=.55,.58,.91 (p<.0001, .001)	CR; CS; CV; D; FA	A
21. Relatives' opinions toward Schizophrenia (ROS)	Magliano et al., 1999 [130]	α*=.56-.66; K**=.36-.84	CS; FA	A
Questionnaire on the Opinions About Mental Illness (QO)	Magliano et al., 2004 [131]	α*=.42-.72; k***=.50-1.0	FA	A
22. EMIC	Chowdhury et al., 2000 [76]	α*=.66-.76; K***=.77-.89		A
23. Stigma Concerns about Mental Health Care (SCAMHC)	Interian et al., 2010 [108]	α*=.69	CS; CV; CR; FA	D
24. Latino Scale for Antidepressant Scale (LSAS)	Interian et al., 2010 [108]	α*=.66	CS; CV; CT; FA	D
25. Devaluation of Consumer Family Scale (DCFS)	Struening et al., 2001 [148]	α*=.82	CV; FA	A
26. Devaluation of consumers scale	Struening et al., 2001 [148]	α*=.71-.77	CV; FA	A
27. Consumer Experiences of Stigma Questionnaire (CESQ)	Bagley & King, 2005 [65]	α*=.79-.82	CC; CS; CR; D; FA	B
28. Attitudes towards Depression and Its Treatment (ATDT)	Gabriel & Violato, 2010 [93]	α*=.57-.79	CT; FA	A
29. Stigmatization Scale (Harvey SS)	Harvey, 2001 [102]	α*=.90, .94	CS; D; FA; CR	B
30. Psychiatric Skepticism Scale (PSS)	Swami & Furnham, 2011 [151]	α*=.92;.94	CS; FA	D
31. Emotional Reactions	Angermeyer & Matschinger, 2003 [63]		CS; FA	A
32. Labeling of mental illness (LMI)	Angermeyer & Matschinger, 2003 [63]	k***=0.85		A

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
33. Personal Attributes	Angermeyer & Matschinger, 2003 [63]		CS; FA	A
34. Depression Attitude Questionnaire (DAQ)	Botega et al., 1992 [69]		FA	A
R-DAQ	Haddad et al., 2015 [101]	α*=.84; ICC**=.62	CT; CS; CV; FA	A
35. Attitudes Toward psychiatry-30 (ATP-30)	Burra et al., 1982 [74]	r it *=.10-.64; Split-half r*=.89, .90; ICC**=.51-.87	CC	D
36. Opening Minds Scale for Health Care Providers (OMS-HC)	Kassam et al., 2012 [116]	α*=.78; .79; .82 r it *=-.13-.57; ICC**=.66 (p<.001)	CS; FA; CT; R	A
37. King Stigma Scale (King SS)	King et al., 2007 [118]	α*=.87 (.64-.87); k**=.41-.71	CS; CC; FA	A
Chinese Stigma Scale (King CSS)	Ho et al., 2015 [106]	α*=.83 (.58-.84)	CC; FA	A
38. Stigma Experiences Scale	Stuart et al., 2005 [149]	α*=.91 KR-20*=.83;	CS	B
39. Attitudes Toward Serious Mental Illness Scale-Adolescent	Watson et al., 2005 [158]		FA	A
40. Self reported prejudiced attitudes	Andersson et al., 2010 [62]	α*=.78	FA	A
41. Self-Stigma of Depression Scale	Barney et al., 2010 [66]	α*=.87 ICC**=.63 (p=.000)	CS; CV; FA	C
42. Employer Attitude Questionnaire (EAQ)	Diksa & Rogers, 1996 [85]		FA	A
43. 15-Item Stigma Questionnaire	Gibbons et al., 2012 [94]	α*=.85; ICC**=.75	CC; CS; CV; CT	B

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
44. Attitudes of Nursing Staff towards Co-Workers Returning from Psychiatric and Physical Illnesses	Glozier et al., 2006 [95]	α*=.76-.88	CS	A
45. Self-Esteem and Stigma Questionnaire (SESQ)	Hayward et al., 2002 [103]	α*=.71-.79; r**=.63 (p<.0001)	CS	A
46. Test of Knowledge About ADHD (KADD)	Hepperlen et al., 2002 [55]	α*=.81-.82;	FA	A
47. Beliefs toward Mental Illness (BMI)	Hirai & Clum, 2000 [105]	α*=.91; r it = (.22< r <.72)	CS; CC; FA	A
48. Depression Self-Stigma Scale (DSSS)	Kanter, 2008 [111]	α*=.79-.95; r it =.44-.83	CS; CC; FA	C
49. General Attitude Questionnaire	Lam et al., 2005 [121]	α*=.88-.93 r**=.72-.94		A
50. Secrecy	Link, 1987 [124]	α*=.73-.83	CS	C
51. Withdrawal	Link, 1987 [124]	α=.73-.83	CS	C
52. Attitudes to Severe Mental Illness (ASMI)	Madianos et al., 2012 [129]	α*=.88(.79-.86); r**=.89-.92 (p<.0001)	CS; P; FA	A
53. Affiliate Self-Stigma Scale (ASSS)	Mak & Cheung et al., 2008 [132]	α*=.94-.95; r it =.51-.81	CS; P; FA	C
Self-Stigma Scale-Short (SSS-S)	Wu et al., 2015 [159]	α*=.95	CC; CS; FA	C
54. Knowledge Test of Mental Illness (KT)	Michaels & Corrigan, 2013 [135]	r**=.50-.70 (p<.05; .001)	CC; CS	A
55. Attitudes Toward Social Competence and Integration of People with Mental Illness	Minnebo & Acker et al., 2004 [136]	α*=.77; .79	FA	A
56. Client Attitude Questionnaire	Morrison & Becker, 1975 [139]	r**=.90; .93		?
57. Libertarian Mental Health Ideology Scale (LMHIS)	Nevid & Morrison, 1980 [141]	α*=.81-.94	CS; FA	D

Measures	Developer/Author	Reliability	Validity & Responsiveness	Content
58. Personal stigma scale	Schneider et al., 2011 [146]	α*=.62-.92	FA	A
59. Child stigma scale	Moses, 2009 [140]	α*=.81 k***=.79-.90	CV; CS; D	C
60. Beliefs and attitudes toward people diagnosed with psychosis	Serra et al., 2013 [58]	α*=.69	FA	A
61. Stigma of Depression Scale	Vega et al., 2010 [154]	α*=.69	FA	A
62. Perceptions of Stigmatization by Others for Seeking Help (PROSH)	Vogel et al., 2009 [156]	α*=.78-.91 r**=.82 (p<.001)	CS; CC; FA	D
63. The Stigma Inventory for Mental Illness	Karidi et al., 2014 [112]	α*=.90 (.75, .85); r**=.80 (p<.001)	CT; CS; CC; FA	C
64. Peer Mental Health Stigmatization Scale	McKeague et al., 2015 [134]	α*=.80 (.70, .75); r**=.65, 75	CT; CS; D; FA	A
65. Endorsed and Anticipated Stigma Inventory (EASI)	Vogt et al., 2014 [157]	r it = .47-.75	CT; CV; D; FA	A & D

1. * internal consistency reliability: Cronbach alpha (α), Kuder-Richardson 20 (KR-20), item to total correlation (r it), split-half reliability; **test-retest reliability: intraclass correlation coefficient (ICC), weighted kappa (k), Pearson correlation coefficient (r), Lin's Pc ; *** inter-rater reliability: weighted kappa (k), Lin'sPc
2. Content validity: CT; Construct validity: CS; Criterion validity: CR; Convergent validity: CV; Concurrent validity: CC; Discriminant validity: D; Factor analysis: FA; Responsiveness: R; Divergent validity: DV; Predictive validity: P.
3. A: Stigma against mental illness or the mentally ill; B: Experienced stigma; C: self-stigma; D: stigma against help-seeking, treatment; mental health institution or psychiatry

Chapter 2 Table 3: Psychometrics of Help-Seeking Measures

Measures	Author/Developer	Reliability	Validity	Content
1. Attitudes towards help-seeking scale (with various modified versions)	Fischer & Turner, 1970 [162]	α*=.83; .86 r it = -.58 - .56 (p<.0001) r**=.73-.89	FA; CS	A/H
Attitudes Toward Seeking Professional Psychological Help Scale (ATSPPH)	Fischer & Farina, 1995 [166]	α*=.77-.98;.84;.90 r it *=.54; r**=.80	FA; CS; CR	
ATSPPH-SF	Elhai, et al., 2008 [165]	α*=.69; .77-.78; .84 r it *>.40 r**= .64 (p=.045)	FA; CS	
2. Intention of Seeking Counseling Inventory (ISCI)	Cepeda-Benito & Short, 1998 [163]	α*=.89	FA; CS	I
3. General Help Seeking Questionnaire (GHSQ)	Deane et al., 2001 [164]	α*=.67, .76, .82	FA	I
	Gulliver et al., 2012 [54]	α*=.57-.77; r**= .42-.91 (p<0.001)		
	Wilson et al., 2005 [178]	α*=.70-.85; ICC**=.86-.92	P; CV; DV	
4. Jorm Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	Jorm, Blewitt et al., 2005 [168]	K***=0.15-1.00		A/T
	Jorm, Mackinnon et al., 2005 [169]		FA	
	Reavley et al., 2014 [154]		CS	
5. Help Seeking Intentions (HSI)	Lee et al., 2014 [170]	α*=.74, .76	FA	I
6. The New Inventory of Attitudes Towards Seeking Mental Health Services (IASMHS)	Mackenzie et al., 2004 [171]	α*=.87 (.76-.82) r**=.64-.91 (p<0.01)	FA; CS	A/H

Measures	Author/Developer	Reliability	Validity	Content
7. Help-Seeking Attitude Scale (HSAS)	Nickerson et al., 1994 [172]	α*=.87	CC	A/H
8. Scale of Attitudes Toward Seeking Psychological Help for Secondary Students (ASPH-S)	Sahin & Uyar, 2011 [175]	α*=.85 (.59-.81) r it *=.41-.57 ICC**=.81	FA	A/H
9. Help Seeking Acceptability (HSA)	Schmeelk-Cone et al., 2012 [176]	α*=.84-.88 r it *=.81-.85	FA; CS	A/H
10. Parental Attitudes Toward Psychological Services Inventory (PATPSI) (based on ATSPPH)	Turner, 2012 [177]	α*=.72-.92 ICC**=.66-.90	FA; CS	A/H

1. * internal consistency reliability: Cronbach alpha (α), item to total correlation (r it); **test-retest reliability: intraclass correlation coefficient (ICC), Pearson correlation coefficient (r); *** inter-rater reliability: weighted kappa (k)
2. Content validity: CT; Construct validity: CS; Criterion validity: CR; Convergent validity: CV; Concurrent validity: CC; Discriminant validity: D; Factor analysis: FA; Divergent validity: DV; Predictive validity: P.
3. A/H: Beliefs/Attitudes towards help-seeking; I: help-seeking intentions; A/T: beliefs/attitudes towards treatment

Chapter 3 Table 1: Study Characteristics

Author/year Measure	ment tool	Description of tool	Population of study	Age of study participants	Study sample size	Country of study	Mental health knowledge type	Psychometric properties of tool assessed
1. Ascher-Svanum & Krause, 1999 [48]	Knowledge about Schizophrenia Questionnaire (KASQ)	25 multiple-choice questions on knowledge of mental illness and management	Inpatients	M=35 (SD=11.4); (range: 18-58)	N=53 (study 1); N=53 (study 2); N=10 (study 3); N=20 (study 4)	US	Schizophrenia	Internal consistency; Reliability; Responsiveness (Sensitivity to change); Content validity
2. Balasubramanian et al., 2013 [217]	Knowledge Questionnaire on Home Care of Schizophrenics (KQHS)	32 item multiple choice questionnaire on four aspects of home care	Home care givers	Unknown	N=21	India	Schizophrenia	Content validity; Internal consistency
3. Compton et al., 2007 [49]	Knowledge about Schizophrenia Test (KAST)	21 multiple choice questions on knowledge of schizophrenia	Community members; Families of people with schizophrenia; police officers; mental health professionals	M=43.7 (SD=12.1) (Community members); M=44.0 (SD=12.8) (families); M=37.8 (SD=7.8) (police officers); M=44.2 (SD=10.1) (mental health professional)	N=144 (community members); N=77 (families members); N=170 (police officers); N=50 (mental health professionals)	US	Schizophrenia	Internal consistency; Construct validity (hypothesis testing); Content validity; Criterion/concurrent validity

Author/year Measure	ment tool	Description of tool	Population of study	Age of study participants	Study sample size	Country of study	Mental health knowledge type	Psychometric properties of tool assessed
4. Compton et al., 2011 [50]	Multiple-Choice Knowledge of Mental Illnesses Test (MC-KOMIT)	33 multiple-choice items on knowledge of common mental illnesses	Police officers	M=38.3 (SD=8.4)	199	US	General knowledge	Internal consistency; Reliability; Construct validity (hypothesis testing); Content validity; Responsiveness
5. Daltio et al., 2015 [218]	Knowledge about Schizophrenia Test (KAST)	17 multiple choice questions on knowledge of schizophrenia	Caregivers of patients with schizophrenia, and patients of other conditions; mental health clinicians	M=56.05 (SD=12.9) (caregivers)	N=89 caregivers of patients with schizophrenia; N=30 caregivers of general patients; N=30 mental health professionals	Portugal	Schizophrenia	Content validity; Cross-cultural validity; Reliability Construct validity
6. Evans-Lacko et al., 2010 [51]	Mental Health Knowledge Schedule (MAKS)	6-point Likert scale on 12 items of stigma knowledge of mental illness	General public	25-45	N=92 (study 1); N=37 (study 2); N=403 (study 3)	UK	General knowledge	Internal consistency; reliability; Content validity
7. Gabriel & Violato, 2009 [52]	Depression Multiple Choice Question (MCQ)	27 multiple-choice items on knowledge of depression	Patients and psychiatrists	M=43 (SD=11.3) (range: 18-65) (patients); M=52 (SD=11.6) (Psychiatrists)	N=63 (patients) N=12 (psychiatrists)	Canada	Depression	Internal consistency; Content validity; Convergent validity; Structural validity (factor analysis)

Author/year Measure	ment tool	Description of tool	Population of study	Age of study participants	Study sample size	Country of study	Mental health knowledge type	Psychometric properties of tool assessed
8. Gulliver et al., 2012 [54]	Depression Literacy (D-Lit)	22 true/false items on knowledge of depression	Elite athletes	M=25.5 (median=24.5) (range: 18-48)	N=40 (study 1); N=12 (study 2)	Australia	Depression	Internal consistency; Reliability
	Anxiety Literacy Questionnaire (A-Lit)						Anxiety	
9. Kiropoulos et al., 2011 [53]	Depression Literacy (D-Lit)	22 true/false items on knowledge about depression	Immigrants	M=65.4 (SD=8.57) (range: 48-88)	202	Australia	Depression	Internal consistency; Reliability
10. Hepperlen et al., 2002 [55]	Test of Knowledge About ADHD (KADD)	22 error-choice items to assess knowledge and attitudes toward students with ADHD	Elementary school teachers	M=39.43 (SD=9.05)	103	US	ADHD	Internal consistency; Content validity; Structural validity (factor analysis);
11. Kronmuller et al., 2008 [56]	Knowledge about Depression and Mania Inventory (KDMI)	44 true/false items on knowledge of Depression and Mania	Patients and relatives	M=45.2 (SD=13.6) (range: 18-82); M=47.4 (SD=14.5) (range: 19-80)	N=112 (patients); N=89 (relatives)	Germany	Depression	Concurrent/criterion validity; Hypothesis testing (Discriminative validity); Content validity; Responsiveness

Author/year Measure	ment tool	Description of tool	Population of study	Age of study participants	Study sample size	Country of study	Mental health knowledge type	Psychometric properties of tool assessed
12. O'Connor & Casey, 2015 [219]	Mental Health Literacy Scale (MHLS)	Multiple choice on 35 items regarding knowledge and attitudes about help-seeking, and ability to recognize disorders	First year university students (S) Mental health professionals (M)	M=21.10±6.27 (S); M=33.09±8.01	372 (S); 43 (M)	Australia	General knowledge	Internal consistency; Reliability; Measurement error; Content validity; Structural validity; Construct validity
13. Pickett-Schenk et al., 2000 [57]	Journey of Hope (JOH) Outcome Survey	4-point Likert scale on 15 items on mental health knowledge	Family members of people with mental illness	M=56.48	424	US	General knowledge	Internal consistency; Construct validity (hypothesis testing); Structural validity (Factor analysis);
14. Serra et al., 2013 [58]	Knowledge of Mental Disorders (KMD)	"Yes, "No", and "I don't know" responses to assess knowledge on the name and characteristics of mental disorders and ability to distinguish them from somatic illnesses	High school students	M=17.3 (SD=1.3); (range: 15-24)	1,023	Italy	General knowledge	Internal consistency; Structural validity (factor analysis); Construct validity (hypothesis testing)

Author/year Measure	ment tool	Description of tool	Population of study	Age of study participants	Study sample size	Country of study	Mental health knowledge type	Psychometric properties of tool assessed
15. Hart et al., 2014 [220]	Adolescent Depression Knowledge Questionnaire (ADKQ)	13 dichotomous and 2 fill-in-the- blank questions on depression knowledge	Grade 9 students	Not reported	8,216	US	Depression	Internal consistency; Structural validity (factor analysis)
16. Swami et al., 2011 [60]	Mental health disorder recognition questionnaire (MDRQ)	7-point Likert scale on 20 statements of mental illness descriptions in which 15 are real and 5 are foils	General public	M=38.11 (SD=14.89)	477	UK	General knowledge	Reliability; Construct validity (hypothesis testing); Convergent validity
17. Wang et al., 2013 [61]	Mental Health Knowledge Questionnaire (MHKQ)	“yes”, and “no” responses to 20 general mental health knowledge questions	Community members	M=50 (SD=17)	1953	China	General knowledge	Internal consistency; Factor analysis

M: mean; SD: standard deviation

Chapter 3 Table 2: Methodological Quality of a Study on Each Measurement Property of a Measurement Tool

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment errors	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
1. Knowledge about Schizophrenia Questionnaire (KASQ) [48]	Ascher-Svanum & Krause, 1999	Poor	Poor	Poor						Poor
2. Knowledge Questionnaire on Home Care of Schizophrenics (KQHS) [217]	Balasubramanian et al., 2013	Poor		Good						
3. Knowledge about Schizophrenia Test (KAST) [49, 218]	Compton et al., 2007	Poor		Excellent			Fair		Fair	
	Daltio et al., 2015		Fair	Excellent				Fair	Fair	
4. Multiple-Choice Knowledge of Mental Illnesses Test (MC-KOMIT) [50]	Compton et al., 2011	Poor	Good	Excellent					Fair	Poor
5. Mental Health Knowledge Schedule (MAKS) [51]	Evans-Lacko et al., 2010	Poor	Fair	Excellent						
6. Depression Multiple Choice Question (DMCQ) [52]	Gabriel & Violato, 2009	Good		Poor		Fair			Fair	
7. Depression Literacy (D-Lit) [53-54]	Gulliver et al., 2012	Poor	Poor							
	Kiropoulos et al., 2011	Poor	Fair							
8. Anxiety Literacy	Gulliver et al.,	Poor	Poor							

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment errors	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
Questionnaire (A-Lit) [54]	2012									
9. Test of Knowledge About ADHD (KADD) [55]	Hepperlen et al., 2002	Good		Poor		Fair				
10. Knowledge about Depression and Mania Inventory (KDMI) [56]	Kronmuller et al., 2008	Poor		Excel-lent			Fair		Fair	Poor
11. Mental Health Literacy Scale (MHLS) [219]	O'Connor & Casey, 2015	Excel-lent	Good	Excel-lent	Good	Excel-lent			Fair	
12. Journey of Hope (JOH) Outcome Survey [57]	Pickett-Schenk et al., 2000	Good				Fair			Fair	
13. Knowledge of Mental Disorders (KMD) [58]	Serra et al., 2013	Good				Fair			Fair	
14. Adolescent Depression Knowledge Questionnaire (ADKQ) [220]	Hart et al., 2014	Good		Poor		Fair				
15. Mental health disorder recognition questionnaire (MDRQ) [60]	Swami et al., 2011		Fair						Fair	
16. Mental Health Knowledge Questionnaire (MHKQ) [61]	Wang et al., 2013	Good							Fair	

Chapter 3 Table 3: Quality of Each Measurement Property

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment error	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
1. Knowledge about Schizophrenia Questionnaire (KASQ) [48]	Ascher-Svanum & Krause, 1999	?	+	?						+
2. Knowledge Questionnaire on Home Care of Schizophrenics (KQHS) [217]	Balasubramanian et al., 2013	?		+						
3. Knowledge about Schizophrenia Test (KAST) [49, 218]	Compton et al., 2007	?		+			-		+	
	Daltio, et al., 2015		-	+				N/A	+	
4. Multiple-Choice Knowledge of Mental Illnesses Test (MC-KOMIT) [50]	Compton et al., 2011	?	-	+					+	+
5. Mental Health Knowledge Schedule (MAKS) [51]	Evans-Lacko et al., 2010	-	+							
6. Depression Multiple Choice Question (DMCQ) [52]	Gabriel & Violato, 2009	-		?		+			+	

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment error	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
7. Depression Literacy (D-Lit) [53-54]	Gulliver et al., 2012	?	-							
	Kiropoulos et al., 2011	?	+							
8. Anxiety Literacy Questionnaire (A-Lit) [54]	Gulliver et al., 2012	?	+							
9. Test of Knowledge About ADHD (KADD) [55]	Hepperlen et al., 2002	+		?		-				
10. Knowledge about Depression and Mania Inventory (KDMI) [56]	Kronmuller et al., 2008	?		+			-		+	+
11. Mental Health Literacy Scale (MHLS) [219]	O'Connor & Casey, 2015	+	+	+	?	?			+	
12. Journey of Hope (JOH) Outcome Survey [57]	Pickett-Schenk et al., 2000	+				?			+	
13. Knowledge of Mental Disorders (KMD) [58]	Serra et al., 2013	-				?			+	

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment error	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
14. Adolescent Depression Knowledge Questionnaire (ADKQ) [220]	Hart et al., 2014	+		?		?				
15. Mental health disorder recognition questionnaire (MDRQ) [60]	Swami et al., 2011		+						?	
16. Mental Health Knowledge Questionnaire (MHKQ) [61]	Wang et al., 2013	-							+	

+: positive rating; -: negative rating: ?: indeterminate rating; N/A: no information provided

Chapter 3 Table 4: Overall Level of Evidence of Measurement Properties

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment errors	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
1. Knowledge about Schizophrenia Questionnaire (KASQ) [48]	Ascher-Svanum & Krause, 1999	x	x	x						x
2. Knowledge Questionnaire on Home Care of Schizophrenics (KQHS) [217]	Balasubramanian et al., 2013	x		++						
3. Knowledge about Schizophrenia Test (KAST) [49, 218]	Compton et al., 2007 Daltio et al., 2015	x	-	+++			-		+	
4. Multiple-Choice Knowledge of Mental Illnesses Test (MC-KOMIT) [50]	Compton et al., 2011	x	--	+++					+	x
5. Mental Health Knowledge Schedule (MAKS) [51]	Evans-Lacko et al., 2010	x	+	+++						
6. Depression Multiple Choice Question (MCQ) [52]	Gabriel & Violato, 2009	--		x		+			+	

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment error	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
7. Depression Literacy (D-Lit) [53-54]	Gulliver et al., 2012 Kiropoulos et al., 2011	x	+/-							
8. Anxiety Literacy Questionnaire (A-Lit) [54]	Gulliver et al., 2012	x	x							
9. Test of Knowledge About ADHD (KADD) [55]	Hepperlen et al., 2002	++		x		-				
10. Knowledge about Depression and Mania Inventory (KDMI) [56]	Kronmuller et al., 2008	x		+++			-		+	x
11. Mental Health Literacy Scale (MHLS) [219]	O'Connor & Casey, 2015	+++	++	+++	x	x			+	
12. Journey of Hope (JOH) Outcome Survey [57]	Pickett-Schenk et al., 2000	++				x			+	
13. Knowledge of Mental Disorders (KMD) [58]	Serra et al., 2013	--				x			+	
14. Adolescent Depression Knowledge Questionnaire (ADKQ) [220]	Hart et al., 2014	++		x		x				

Measurement tool	Study Author	Internal consis-tency	Reliability	Content validity	Measure-ment error	Structural validity	Criterion validity	Cultural validity	Hypothesis testing	Responsive-ness
15. Mental health disorder recognition questionnaire (MDRQ) [60]	Swami et al., 2011		+						+	
16. Mental Health Knowledge Questionnaire (MHKQ) [61]	Wang et al.,2013	++							+	

Note: +++ or --- = strong evidence; ++ or-- = moderate; + or- = limited evidence; +/-: conflict findings; x=studies of poor methodological quality or studies with indeterminate property quality

Chapter 4 Table 1: Study Characteristics

Measurement tools	Response options	Population	Age	Sample	Country	Content
1. Self reported prejudiced attitudes (SRPA) [62]	9 4-point scaled items	Upper secondary schools	16-19	4,046	Norway	A
2. Personal attributes (PA) [63]	8 5-point scaled items	Community members	≥18	5,025	German	A
3. Labeling of mental illness (LMI) [63]	Open-ended questions for a vignette	Community members	≥18	5,025	German	A
4. Peer Mental Health Stigmatization scale (PMHSS) [134]	24 5-point scaled items	Children and adolescents in schools	M=12.99±1.6	562	Ireland	A
5. Knowledge Test of Mental Illness (KT) [135]	14 items with numerical or true/false responses	College students, community members, mental health providers and consumers	M=21.6±3.2; 33.1±7.4; 45.5±11.4; 45.4±11.2	35; 203; 133; 74	US	A
6. Day's Mental Illness Scale (DMIS) [84]	28 7-point scaled items	College students, community members and people with mental illness	M=24.84; 18.60; 45	341; 42; 20	US	A
7. EMIC [76]	13 4-point scaled items	Laypersons and health care providers	?	38	India	A
8. Employer Attitudes towards Mental Illness questionnaire (EAQ) [85]	38 5-point scaled items	Employers	?	373	US	A
9. Attitudes of Nursing Staff towards Co-Workers Returning from Psychiatric and Physical Illnesses (ANCW) [95]	12 4-point scaled items	Nursing staff	M=35.6-38.6	117	UK	A
10. Depression Stigma scale (DSS)/PPSM [98]	18 5-point scaled items	Community members	M=36.4±9.4	525	Australia	A
DSS [99]	18 5-point scaled items	Community members	Median=45-49; M=35.9±9.2; 35.3±8.76	1,001; 5572; 487	Australia	A
11. DSS revised [54]	9 5-point scaled items	Elite athletes	M=25.5	59	Australia	A
12. Generalized Anxiety Stigma Scale (GASS) [100]	20 5-point scaled items	General public	M=46.6±13.25	617; 212	Australia	A
13. GASS revised [54]	10 5-point scaled items	Elite athletes	M=25.5	59	Australia	A

Measurement tools	Response options	Population	Age	Sample	Country	Content
14. Test of Knowledge About ADHD (KADD) [55]	22 error-choice items	Elementary school teachers	M=39.43	103	US	A
15. Beliefs toward Mental Illness (BMI) [105]	24 6-point scaled items	College students	M=25.3±5.1	216	US	A
16. Opening Minds Scale for Health Care Providers (OMS-HC) [116]	20 5-point scaled items	Health care providers	≥18	787	Canada	A
OMS-HC [137]	20 5-point scaled items	Health care providers	18-65	1,523	Canada	A
17. ADHD Stigma Questionnaire (ASQ) [117]	26 4-point scaled items	Adolescents at risk of ADHD	M=15.6±1.8	301	US	A
ASQ [67]	26 4-point scaled items	Teachers	M=42.32±12.61	268	US	A
18. ADHD Stigma (ADHD S) [90]	37 6-point scaled items	College students, community members, teaches and physicians	M=31.3±14.8; M=50.6-52.3	1033; 228	Netherlands	A
19. King SS Chinese version (King CSS) [106]	14 5-point scaled items	People with mental illness	M=51.2±11.34	114	China	A
20. Self-Esteem and Stigma Questionnaire (SESQ) [103]	8 6-point scaled items	Patients with depression	M=43±11	186	UK	A
21. Devaluation-Discrimination tool (DD) [124]	12 4-point scaled items	Community members and people with mental illness	M=32.71-40.29	593	US	A
Perceived DD [73]	12 4-point scaled items	Patients with mental illness	M=46	40	Sweden	A
Perceived DD (depression) [108]	12 4-point scaled items	Latino primary care patients	≥18	200	US	A
22. Depression is a Matter of Will (DMW) [64]	16 4-point scaled items	Community members	M=46.9±17.3	5,520	Finland	A
23. Public stigma (PS) [140]	14 4-point scales items	Youth with mental illness	12-18	60	US	A
24. Perceived dangerousness (PD) [125]	8 6-point scaled items	Community members	M=47.6	152	US	A
25. PD/Link Stigma scale [65]	11 5-point scaled items	Patients with mental illness	M=42.2	83	UK	A
26. Dangerousness Scale (DS) [143]	8 7-point scaled items	College students	?	329	US	A
27. Social distance (SD) [125]	7 4-point scaled items	Community members	M=47.6	152	US	A
SD [143]	7 4-point scaled items	College students	?	329	US	A
28. SD revised [108]	6 multiple choice items	Latino primary care patients	≥18	200	US	A
29. Reported and Intended Behaviour Scale (RIBS) [86]	8 items	General public	M=46±18.6	6,954	UK	A

Measurement tools	Response options	Population	Age	Sample	Country	Content
RIBS [87]	4 items on an ordinal scale and 4 items on multiple choices	General public	M=38.1±13.4; 36.9±14.1	403; 83	UK	A
RIBS [88]	4 items on an ordinal scale and 4 items on multiple choices	General public	25-45	92; 37; 403	UK	A
RIBS [89]	4 items on an ordinal scale and 4 items on multiple choices	Medical students	M=23.5	1,452	UK	A
RIBS Japanese version [160]	4 items on an ordinal scale and 4 items on multiple choices	Undergraduate and graduate students	M=22.61±2.47	224	Japan	A
30. Social Contact Scale (SCS) [110]	8 4-point scaled items	Secondary school students	M=13.3±1.26	1,223	Jamaica	A
31. Social Supports Acceptance Scale (SSAS) [133]	? 4-point scaled items	People with mental illness	18-70	70	US	A
32. Attitudes to Mental Illness Questionnaire (AMID) [127]	5 5-point scaled items	General public	M=46.3±15.7	1,079	UK	A
33. Attitudes to Severe Mental Illness (ASMI) [129]	35 6-point scaled items	General public	M=41.5±10.61; 43.71±11.18	2,039	Greece	A
34. Attitudes Toward Social Competence and Integration of People with Mental Illness (ASCI) [136]	8 5-point scaled items	Belgian high school students	M=16.8±1.6	207	Belgium	A
35. Client Attitude Questionnaire (CAQ) [139]	20 true/false/don't know items	Psychiatric professionals	?	13	US	A
36. Beliefs and attitudes toward people diagnosed with psychosis (BAP) [58]	6 5-point scaled items	High school students	M=17.3±1.4; 17.3±1.3	1,023	Italy	A
37. Devaluation of consumers family scale (DCFS) [148]	15 4-point scaled items	Caregivers of people with mental illness	M=50±14.3	461	US	A
Stigma-Devaluation scale [83]	15 4-point scaled items	Family members of people with mental illness	M=44.5±11.7	164	Jordan	A
38. Adolescent Attitudes toward Serious Mental Illness (ATSMI-AV) [158]	24 5-point scaled items	High school students	Grades 9-12	415	US	A

Measurement tools	Response options	Population	Age	Sample	Country	Content
39. Attitudes towards Acute Mental Health Care (ATAMHS) [233]	33 7-point scaled items and semantic differentials	Nurses	35-39	140	UK	B
40. Attitudes towards Psychiatry 30 (ATP-30) [74]	30 5-point scaled items	Medical students and residents	?	189	Canada	B
41. Latino Scale for Antidepressant Stigma (LSAS) [108]	7 3-point scaled items	Latino primary care patients	≥18	200	US	B
42. Stigma Concerns about Mental Health Care (SCMHC) [108]	3 scaled items	Latino primary care patients	≥18	200	US	B
43. Stigma Scale for Receiving Psychological Help (SSRPH) [120]	5 3-point scaled items	College students	M=18.4±1.32	311	US	B
44. Psychiatric Skepticism Scale (PSS) [151]	16 7-point scaled items	General public	M=37.55±14.67	564	UK	B
45. Perceptions of Stigmatization by Others for Seeking Help (PSOSH) [156]	21 5-point scaled items	College students	?	985; 842; 506; 144; 130	US	B
46. British Omnibus National Survey (ONS) [119]	11 5-point scaled items	Community members	≥18	5,251	US	A & B
47. Changing Mind (CM) [150]	56 5-point scaled items	Nursing students	20-25	51	Sweden	A & B
48. Mental Illness: Clinician's Attitudes (MICA) [113]	16 6-point scaled items	Medical students, psychiatry trainees,	M=22.4-22.9	23-188	UK	A & B
MICA version 4 [92]	16 6-point scaled items	Nursing students	M=25.56±7.29	191	UK	A & B
49. Libertarian Mental Health Ideology scale (LMHIS) [141]	39 5-point scaled items	Mental health professionals and students	M=34	227	US	A & B
50. Depression Attitude Questionnaire (DAQ) [69]	20 items on a 100 mm visual analogue scale	General practitioners	M=41±7.4	72	UK	A & B
DAQ [237]	20 items on a 100 mm visual analogue scale	Nurses and home care staff	M=44.7±9.3	189	UK	A & B
51. DAQ revised [245]	24 5-point scaled items	Pharmacists	M=45.2±11.1	200	Belgium	A & B
52. R-DAQ [101]	22 5-point scaled items	Health care providers	?	1,193	UK	A & B
53. Opinions about Mental Illness (OMI) [78]	70 6-point scaled items	Health care providers and hospital staff	?	1,194	US	A & B

Measurement tools	Response options	Population	Age	Sample	Country	Content
OMI [241]	51 6-point scaled items	Community members	M=40.9±13.1	1,574	Greece	A & B
OMI [147]	51 6-point scaled items	Health care providers and hospital staff	?	1,200	US	A & B
54. OMI Chinese [142]	45 6-point scaled items	Secondary students	M=15.04±1.18	117	China	A & B
55. Community Attitudes towards Mental Illness (CAMI) [153]	40 9-point scaled items	Community members	?	1,090	Canada	A & B
CAMI [96]	40 9-point scaled items	Undergraduate students	18-40	102	US	A & B
CAMI [97]	40 9-point scaled items	Undergraduate students	M=20.54±2.30	53	US	A & B
CAMI [104]	40 9-point scaled items	Undergraduate students	18-30	86	US	A & B
CAMI [138]	40 9-point scaled items	Nurses	M=40±10	858	6 European countries	A & B
CAMI Chinese [244]	40 9-point scaled items	Mental health professionals	?	100	China	A & B
CAMI [250]	40 9-point scaled items	General public	≥18	192	UK	A & B
56. CAMI revised [234]	31 5-point scaled items	Community members	≥15	2,000	Canada	A & B
57. Mental Health Attitude Survey for Police (MHASP) [77]	37 3/4-point scaled items	Police officers	M=41.34±9.09	394	US	A & B
58. CAMI Swedish [107]	20 6-point scaled items	Student nurses	M=27.9±7.5	256	Sweden	A & B
59. CAMI/FABI (20 item) [150]	20 5-point scaled items	Nursing students	20-25	51	Sweden	A & B
60. Relatives' opinions toward Schizophrenia (ROS) [130]	28 10-point scaled items	Relatives of people with mental illness	M=55.9±14.8	103	Italy	A & B
61. R-AQ [144]	9 7-point scaled items	High school students	M=20.15±6.33; M=20.50±5.87	210	US	A & B
62. Attitudes towards Depression and Its Treatment (ATDT) [93]	27 5-point scaled items	Patients with depression; mental health experts	M=43; 52±11.6	63; 12	Canada	A & B
63. ATDT revised [109]	25 5-point scaled items	Community members	M=32.2±12.9	203	Australia	A & B
64. General Attitude Questionnaire (GAQ) [121]	5 items measured on 0 to 100 visual-analogue scale	Community members	M=41.35	110	UK	A & B
65. Endorsed and Anticipated Stigma Inventory (EASI) [157]	40 5-point scaled items	Military personnel and veterans	M=37.52±9.99	702	US	A & B
66. emotional reactions (ER) [63]	9 5-point scaled items	Community members	≥18	5,025	German	C
67. Affective Reaction Scale (ARS) [143]	10 7-point scaled items	College students	?	329	US	C

Measurement tools	Response options	Population	Age	Sample	Country	Content
68. Attribution questionnaire (AQ) [79]	27 9-point scaled items	College students	M=26.3±12.2	213	US	A & C
AQ [72]	27 9-point scaled items	College students	M=19.2±	774	US	A & C
AQ [80]	27 9-point scaled items	College students	M=25.33±8.77	518	US	A & C
AQ [236]	27 9-point scaled items	College students	M=25.7±9.5	54	US	A & C
AQ-27-Italian [126]	27 9-point scaled items	Relatives of college students	M=40.15±16.36	214	Italy	A & C
69. Consumer Experiences of Stigma Questionnaire (CESQ) -7 items [65]	7 5-point scaled items	Patients with mental illness	M=42.2	83	UK	D
70. CESQ – 9 items [60]	7 5-point scaled items	Patients with mental illness	M=41.5	509	Poland	D
71. Rejection experience (RE) –Swedish [73]	11 5-point scaled items	Patients with mental illness	M=46	40	Sweden	D
72. Personal rejection scale/RE (PRS) [140]	6 yes/no questions	Youth with mental illness	12-18	60	US	D
73. Discrimination and Stigma Scale (DISC) [71]	22 4-point scaled items	Patients with mental illness	M=41.2±10.9	86	UK	D
74. DISC revised [249]	32 7-point scaled items plus 4 interview questions	People with schizophrenia	M=39.2±11.32	732	27 countries	D
75. Questionnaire on Anticipated Discrimination (QUAD) [91]	17 4-point scaled items	People with mental illness	M=54±12.69	117	UK	D
76. 15 item Stigma Questionnaire (SQ-15) [94]	15 5-point scaled items	People with mental illness	M=45.7±12; 46.9±16.7	89; 33	Canada	D
77. Harvey stigma scale (Harvey SS) [102]	18 5-point scaled items	College students	M=24.07±7.34	197	US	D
78. Harvey SS revised [65]	15 five-point scaled items	Patients with mental illness	M=42.2	83	UK	D
79. Link's Rejection experiences (RE) [240]	12 multiple choice items	Men with mental illness	M=34	84	US	D
80. Stigmatizing experiences scale (SES) [149]	17 items on prevalence and frequency of stigma experience	People with mental illness	20-79; median=46	88	Canada	D
81. Self-stigma of depression scale (SSDS) [66]	16 5-point scaled items	Community members	M=50.9	1,312	Australia	E

Measurement tools	Response options	Population	Age	Sample	Country	Content
82. Self-stigma of mental illness (SSMI) [82]	60 9-point scaled items	People with mental illness	M=41.8±9.6; 44.5±8.5	54; 60	US	E
83. SSMI short form [81]	20 9-point scaled items	People with mental illness	M=44.5; 27.8; 35.1; 44.8	71; 60; 30; 85	US, German, Switzerland	E
84. Depression Self-Stigma Scale (DSSS) [111]	32 7-point scaled items	Undergraduates and community members	M=20.93±3.38; 38±13.76	391	US	E
85. The Stigma Inventory for Mental Illness (SIMI) [112]	12 5-point scaled items	Patients with schizophrenia	M=39.7±9.4	100	Greece	E
86. Link's Secrecy [238]	8 6-point scaled items	Community members and people with mental illness	M=32.71-40.29	429; 164	US	E
87. Link's Secrecy [239]	5 6-point scaled items	People with mental illness	?	152	US	E
88. Link's Withdrawal [238]	4 6-point scaled items	Community members and people with mental illness	M=32.71-40.29	429; 164	US	E
89. Link's Withdrawal [239]	7 6-point scaled items	People with mental illness	?	152	US	E
90. Affiliate self-stigma scale (ASSS) [132]	22 4-point scaled items	Caregivers of people with intellectual disability and mental illness	M=42.81±5.41; 54.21±13.20	210; 108	China	E
91. Self-Stigma Scale (SSS) [242]	9 4-point scaled items	People with mental illness, recent immigrants	M=40.07±10.16; 33.98±6.31;	175; 110;	China	E
Self-Stigma Scale-Short (SSS-S) [159]	9 4-point scaled items	People with mental illness	M=40.53±10.38; 46.52±11.29	161; 189	China	E
92. Child stigma scale (CSS)/self-stigma [140]	5 4-point Likert scaled items	Youth with mental illness	12-18	60	US	E
93. Secrecy scale (Secrecy S) [140]	7 items	Youth with mental illness	12-18	60	US	E
94. Internalized stigma of mental illness (ISMI) [145]	29 4-point scaled items	People with mental illness	M=49.5±8.7	127	US	E
ISMI [235]	29 4-point scaled items	People with bipolar disorder or depression	M=45.67 (SD=12.81)	1,182	13 European countries	E
ISMI [243]	29 4-point scaled items	People with mental illness	M=51±10	82	US	E
ISMI Chinese [75]	29 4-point scaled items	People with mental illness	M=43.76±11.27	347	China	E
ISMI Arabic [115]	29 4-point scaled items	Arab refugees with mental illness	M=39.66±11.45	330	US	E

Measurement tools	Response options	Population	Age	Sample	Country	Content
ISMI [122]	29 4-point scaled items	People with mental illness	M=43.6±11.76	160	China	E
ISMI [246]	29 4-point scaled items	People with schizophrenia	M=37.3±11.9	157	Austria	E
ISMI [247]	29 4-point scaled items	People with schizophrenia	M=37.3±11.9	157	Austria	C
ISMI [248]	29 4-point scaled items	Members of depression and anxiety organization	M=37±11.3	142	South Africa	E
95. ISMI revised [232]	24 4-point scaled items	People with schizophrenia	M=33.3±8.9	212	Ethiopia	E
96. ISMI short [70]	10 4-point scaled items	People with mental illness	M=49.5; 49.6	127; 760	US	E
97. ISMI (Parent) [161]	17 4-point scaled items	Parents of people with mental illness	M=58.46±4.71	194	Israel	E
98. Self Stigma of Seeking Psychological Help (SSOSH) [155]	10 5-point scaled items	College students	?	583; 470; 546; 217; 655	US	E
99. Personal stigma scale (Personal SS) [146]	26 5-point scaled items	People with mental illness	?	243	UK	D & E
100. Stigma of Depression scale (SODS) [154]	7 items	Latino people with depression	M=50.6±11.3	200	US	A & E
101. King Stigma Scale (King SS) [118]	28 5-point scaled items	Patients with mental illness	M=42.9±12.4	109	UK	A, C, D

A: Stigma against mental illness or the mentally ill; *B:* stigma against help-seeking, treatment, mental health institution or psychiatry; *C:* Emotional responses to mental illness; *D:* Experienced stigma; E: self-stigma; ?: not reported

Chapter 4 Table 2: Methodological Quality of Included Studies, Quality of Each Measurement Property, and Overall Level of Evidence of Measurement Properties

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
1. Self reported prejudiced attitudes (SRPA) [62]	G; +; M(++)			F; ?; U(x)				
2. Personal attributes (PA) [63]				F; +; L(+)	F; ?; U (x)			
3. Labeling of mental illness (LMI) [63]	P; +; U(x)				F; +; L(+)			
4. Peer Mental Health Stigmatization scale (PMHSS) [134]	G; +; M(++)	F; -; L(-)	P; -; U(x)	F; +; L(+)	F; +; L(+)			
5. ??Knowledge Test of Mental Illness (KT) [135]		F; +/-; C(+/-)		F; -; L(-)			F; -; L(-)	
6. Day's Mental Illness Scale (DMIS) [84]	G; +; M(++)		P; ?; U (x)	F; +; L(+)	F; +; L(+)			
7. ??EMIC [76]	P; +; U (x)	P; +; U (x)						
8. **Employer Attitudes towards Mental Illness questionnaire (EAQ) [85]	G; +; M(++)		F; +; L(+)	F; +; L(+)				
9. Attitudes of Nursing Staff towards Co-Workers Returning from Psychiatric and Physical Illnesses (ANCW) [95]	P; +; U (x)		F; +; L(+)		F; +; L(+)			
10. Depression Stigma scale (DSS)/PPSM [98]	P; +; U (x)	F; +; L(+)						
DSS [99]	G; +; U (x)			F; -; L(-)	F; +; L(+)			
11. ??DSS revised [54]	P; +; U (x)	P; +; U (x)						
12. Generalized Anxiety Stigma Scale (GASS) [100]	E; +; S(+++)	G; -; M(--)		E; +; S(+++)	E; +; S(+++)			
13. ??GASS revised [54]	P; +; U (x)	P; +; U (x)						
14. Test of Knowledge About ADHD (KADD) [55]	G; +; M(++)			F; -; L(-)				
15. **Beliefs toward Mental Illness (BMI) [105]	G; +; M(++)			F; +; L(+)	F; +; L(+)			

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
16. Opening Minds Scale for Health Care Providers (OMS-HC) [116]	G; +; M(++)	F; -; L(-)	E; +; S(+++)	F; -; L(-)				
OMS-HC [137]	G; +; M(++)			F; -; L(-)				P; ?; U (x)
17. ADHD Stigma Questionnaire (ASQ) [117]	G; +; M(++)	F; +/-; C(+/-)		F; ?; U(x)	F; +; U(x)			
ASQ [67]	G; +; M(++)			F; ?; U(x)	F; ?; U(x)			
18. ADHD stigma (ADHD S) [90]	G; +; M(++)			F; +; L(+)	F; ?; U(x)			
19. King SS Chinese version (King CSS) [106]	G; +; M(++)			F; +; L(+)			F; -; L(-)	
20. Self-Esteem and Stigma Questionnaire (SESQ) [103]	F; +; L(+)	F; -; L(-)		F; ?; U (x)	F; -; L(-)			
21. Devaluation-Discrimination tool (DD) [124]	P; +; U(x)				F; +; C(+/-)			
Perceived DD [73]	P; +; U(x)				F; ?; C(+/-)			
Perceived DD (depression) [108]	G; -; U(x)			F; +; L(+)	F; -; C(+/-)			
22. Depression is a Matter of Will (DMW) [64]	G; -; M(--)			F; -; L(-)	F; +; L(+)			
23. Public stigma (PS) [140]	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)			
24. Perceived dangerousness (PD) [125]	P; +; U(x)				F; +; L(+)			
25. PD/Link Stigma scale [65]	P; +; U(x)				F; +; L(+)			
26. ??Dangerousness Scale (DS) [143]	P; +; U(x)				F; -; L(-)			
27. ??Social distance (SD) [125]	P; +; U(x)				F; +; C(+/-)			
SD [143]	P; +; U(x)				F; -; C(+/-)			
28. Social distance revised [108]	G; +; M(++)			F; +; L(+)	F; -; L(-)			
29. Reported and Intended Behaviour Scale (RIBS) [86]	P; +; U(x)	F; +; L(+)						
RIBS [87]	P; +; U(x)	F; +; L(+)						
RIBS [88]	P; +; U(x)	F; +; L(+)	G; +; M(++)					

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
RIBS [89]	P; +; U(x)	F; +; L(+)						
RIBS Japanese version [160]	G; +; U(x)	P; +; L(+)		F; ?; U(x)	F; +; L(+)			
30. ??Social Contact Scale (SCS) [110]	G; +/-; C(+/-)			F; ?; U(x)				
31. Social Supports Acceptance Scale (SSAS) [133]	P; +; U(x)				F; +; L(+)			
32. **Attitudes to Mental Illness Questionnaire (AMID) [127]		F; +; L(+)		F; +; L(+)				
33. **Attitudes to Severe Mental Illness (ASMI) [129]	G; +; M(++)	F; +; L(+)	F; +; L(+)	F; +; L(+)	F; +; L(+)			
34. Attitudes Toward Social Competence and Integration of People with Mental Illness (ASCI) [136]	G; +; M(++)			P; ?; U(x)				
35. ??Client Attitude Questionnaire (CAQ) [139]		P; +; U(x)						
36. Beliefs and attitudes toward people diagnosed with psychosis (BAP) [58]	G; +; M(++)			F; ?; U(x)				
37. **Devaluation of consumers family scale (DCFS) [148]	G; +; M(++)			F; +; L(+)	F; +; L(+)			
Stigma-Devaluation scale [83]	E; +; M(++)		F; +; L(+)	E; ?; L(+)		P; N/A; U(x)		
38. **Adolescent Attitudes toward Serious Mental Illness (ATSMI=AV) [158]				G; +; M(++)				
39. Attitudes towards Acute Mental Health Care (ATAMHS) [233]	F; +; L(+)			F; -; L(-)				
40. Attitudes towards Psychiatry 30 (ATP-30) [74]	F; ?; U (x)	F; +/-; C(+/-)	P; ?; U (x)	F; -; L(-)	F; +; L(+)			
41. Latino Scale for Antidepressant Stigma (LSAS) [108]	G; -; M(--)			F; +; L(+)	F; -; L(-)			
42. Stigma Concerns about Mental Health Care (SCMHC) [108]	G; +/-; C(+/-)			F; +; L(+)	F; -; L(-)			

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
43. **Stigma Scale for Receiving Psychological Help (SSRPH) [120]	G; +; M(++)			F; +; L(+)	F; +; L(+)			
44. **Psychiatric Skepticism Scale (PSS) [151]	G; +; M(++)			F; +; L(+)	F; +; L(+)			
45. Perceptions of Stigmatization by Others for Seeking Help (PSOSH) [156]	G; +; M(++)	F; +; L(+)	P; ?; U(x)	F; +; L(+)	F; +; L(+)		F; +; L(+)	
46. ??British Omnibus National Survey (ONS) [119]	G; -; M(--)			F; -; L(-)	F; -; L(-)			
47. ??Changing mind (CM) [150]	P; -; U(x)	F; -; L(-)						
48. Mental Illness: Clinician's Attitudes (MICA) [113]	G; +; M(++)	F; +; L(+)	G; +; L(+)	F; +; L(+)	F; +/-; C(+/-)			P; ?; U (x)
MICA version 4 [92]	G; +; M(++)		F; +; L(+)	F; +; L(+)	F; -; C(+/-)			
49. Libertarian Mental Health Ideology scale (LMHIS) [141]	G; +; M(++)	F; +; L(+)	G; +; M(++)	F; -; L(-)	F; +; L(+)			
50. ??Depression Attitude Questionnaire (DAQ) [69]				F; -; L(-)				
DAQ [237]				F; -; L(-)	F; -; L(-)			
51. ??DAQ revised [245]	F; +/-; C(+/-)			F; -; L(-)				
52. R-DAQ [101]	G; +; M(++)	F; -; L(-)	G; +; M(++)	F; -; L(-)				
53. Opinions about Mental Illness (OMI) [78]				F; +; L(+)	F; ?; U (x)			
OMI [241]			F; +; L(+)	F; +; L(+)				
OMI [147]	G; +/-; C(+/-)			F ?; L(+)				
54. ??OMI Chinese [142]	P; +; U (x)			F; ?; U (x)				
55. Community Attitudes towards Mental Illness (CAMI) [153]	G; +/-; C(+/-)		P; ?; U(x)	F; ?; C(+/-)	F; +; L(+)			
CAMI [96]	P; +/-; C(+/-)				P; +/-; L(+)			
CAMI [97]	P; +/-; C(+/-)				P; +; L(+)			
CAMI [104]	G; +/-; C(+/-)				F; +; L(+)			
CAMI [138]				F; ?; C(+/-)				

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
CAMI Chinese [244]				F: -; C(+/-)	F: +; L(+)			
CAMI [250]				F: -; C(+/-)	F; +; L(+)			
56. **CAMI revised [234]				F; +; L(+)	F; +; L(+)			
57. Mental Health Attitude Survey for Police (MHASP) [77]	G; +; M(++)			F; -; L(-)	P; +; U(x)			
58. CAMI Swedish [107]	G; +; M(++)			F; ?; U(x)		F; N/A; U(x)		
59. ??CAMI/FABI (20 item) [150]	P +; U(x)	F; -; L(-)						
60. Relatives' opinions toward Schizophrenia (ROS) [130]	G; -; M(--)	F; +/-; C(+/-)		F; +; L(+)	P; -; U(x)			
61. R-AQ [144]	G; +; M(++)			F; -; L(-)				
62. Attitudes towards Depression and Its Treatment (ATDT) [93]	G; +; M(++)		P; ?; U (x)	F; +; L(+)				
63. ??ATDT revised [109]	G; -; M(--)			F; ?; U (x)				
64. ??General Attitude Questionnaire (GAQ) [121]	P; +; U(x)	F; +/-; C(+/-)						
65. Endorsed and Anticipated Stigma Inventory (EASI) [157]	F; ?; U(x)		P; ?; U(x)	F; ?; U(x)	F; +; L(+)			
66. Emotional reactions (ER) [63]				F; +; L(+)	F; ?; U (x)			
67. ??Affective Reaction Scale (ARS) [143]	P; +; U(x)				F; -; L(-)			
68. Attribution questionnaire (AQ)-27 [79]				F; ?; C(+/-)	F; ?; C(+/-)			
AQ [72]	G; +/-; C(+/-)	F; +; L(+)		F; +; C(+/-)	F; +/-; C(+/-)			
AQ [80]	P; +; C(+/-)				F; +; C(+/-)			
AQ [236]		F; +/-; L(+)						
AQ-27-Italian [126]	G; +; C(+/-)	F; +; L(+)		F; ?; C(+/-)	F; +; C(+/-)	F; N/A; U (x)		
69. Consumer Experiences of Stigma Questionnaire (CESQ) -7 items [65]	P; +; U(x)				F; +; L(+)			
70. CESQ – 9 items [60]	G; +; M(++)			F; ?; U (x)	F; +; L(+)	F; N/A; U (x)		

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
71. ??Rejection experience (RE) –Swedish [73]	P; +; U(x)				F; ?; U(x)			
72. Personal rejection scale/RE (PRS) [140]	G; +; M(++)	F; +; L(+)		F; ?; U (x)	F; +; L(+)			
73. ??Discrimination and Stigma Scale (DISC) [71]	P; +; U(x)	F; +/-; C(+/-)			F; -; L(-)			
74. **DISC revised [249]			E; +; S(+++)					
75. Questionnaire on Anticipated Discrimination (QUAD) [91]	G; +; M(++)	F; +; L(+)		F; -; L(-)	F; -; L(-)			
76. 15 item Stigma Questionnaire (SQ-15) [94]	P; +; U (x)	F; +; L(+)	G; ?; U (x)		F; +; L(+)		F; +; L(+)	
77. Harvey stigma scale (Harvey SS) [102]	G; +; M(++)		F; +; L(+)	F; ?; U (x)	F; +; L(+)			
78. Harvey SS revised [65]	P; +; U (x)				F; +; L(+)			
79. Link's Rejection experiences (RE) [240]	G; +; M(++)			F; ?; U(x)				
80. Stigmatizing experiences scale (SES) [149]	P; +; U(x)		G; +; M(++)		F; +; L(+)			
81. Self-stigma of depression scale (SSDS) [66]	G; +; M(++)	F; -; L(-)		F; +; L(+)	F; +; L(+)			
82. Self-stigma of mental illness (SSMI) [82]	P; +/-; U (x)	F; +/-; C(+/-)			F; +; L(+)			
83. SSMI short form [81]	P; +; U (x)				F; +; L(+)			
84. Depression Self-Stigma Scale (DSSS) [111]	G; +; M(++)			F; ?; U (x)	F; +; L(+)			
85. The Stigma Inventory for Mental Illness (SIMI) [112]	G; +; M(++)	F; +; L(+)	P; ?; U (x)	F; ?; U (x)	P; -; U (x)		F; -; L(-)	
86. Link's Secrecy [238]	G; +; M(++)			F; ?; U(x)	F; +; L(+)			
87. Link's Secrecy [239]	G; +; M(++)			F; ?; U(x)	F; +; L(+)			
88. Link's Withdrawal [238]	G; -; M(--)			F; ?; U(x)	F; +; L(+)			
89. Link's Withdrawal [239]	G; -; M(--)			F; ?; U(x)	F; +; L(+)			
90. Affiliate self-stigma scale (ASSS) [132]	G; +; M(++)			F; -; L(-)	F; +; L(+)			

Measurement tools	Internal consistency	Reliability	Content validity	Structural validity	Hypothesis testing (construct validity)	Cross-cultural validity	Criterion validity	Respon-siveness
91. Self-Stigma Scale (SSS) [242]	G; +; M(++)		E; +; S(+++)	F; +; C(+/-)	F; +; C(+/-)			
Self-Stigma Scale-Short (SSS-S) [159]	G; +; M(++)			F; ?; C(+/-)	F; -; C(+/-)			
92. Child stigma scale (CSS)/self-stigma [140]	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)			
93. Secrecy scale (Secrecy S) [140]	G; +; M(++)	F; +; L(+)		F; ?; U(x)	F; +; L(+)			
94. Internalized stigma of mental illness (ISMI) [145]	G; +; M(++)	P; +; C(+/-)		F; ?; L(+)	F; +; L(+)			
ISMI [235]	P; +; M(++)				F; +; L(+)			
ISMI [243]					F; +; L(+)			
ISMI Chinese [75]	G; +; M(++)	F; +; C(+/-)		F; ?; L(+)	F; +/-; L(+)			
ISMI Arabic [115]	G; +; M(++)			F; +; L(+)	F; -; L(+)		F; -; L(-)	
ISMI [122]	G; +; M(++)	F; +/-; C(+/-)		F; +; L(+)	F; +; L(+)			
ISMI [246]					G; +; L(+)			
ISMI [247]				F; +; L(+)	F; +; L(+)			
ISMI [248]					F; +; L(+)			
95. **ISMI revised [232]	G; +; M(++)			F; +; L(+)	F; +; L(+)			
96. ISMI short [70]	P; +; U(x)		F; +; L(+)		F; +; L(+)			
97. ISMI (Parent) [161]	G; +; M(++)		P; ?; U(x)	F; +; L(+)				
98. Self Stigma of Seeking Psychological Help (SSOSH) [155]	G; +; M(++)	F; +; L(+)	P; ?; U(x)	F; +; L(+)	F; +; L(+)		F; -; L(-)	
99. **Personal stigma scale (Personal SS) [146]	G; +; M(++)			F; +; L(+)				
100. ??Stigma of Depression scale (SODS) [154]	G; -; M(--)			F; ?; U(x)				
101. King Stigma Scale (King SS) [118]	G; +; M(++)	F; +/-; C(+/-)		F; +; L(+)	F; +; L(+)			

Study quality: E=Excellent, G=Good, F=Fair, P=Poor;

Quality of each measurement property: positive rating (+), negative rating (-), indeterminate rating (?), conflicting rating (+/-);

Overall level of evidence: Strong (S) (+++ or ---), Moderate (M) (++ or --), Limited (L) (+ or -), Conflicting (C) (+/-), or unknown (U) (x); N/A=Not applicable.

**: 12 tools of which all their measurement properties met the criteria of Limited (+ or -) (minimum acceptable) evidence or above;

??: 20 tools of which no measurement properties met the criteria of minimum acceptable evidence (limited level of evidence) or above.

Chapter 5 Table 1: Study Characteristics

Author/year Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
1. Ang, et al., 2007 [253] Attitudes toward Seeking Professional Psychological Help Scale (ATSPPH)	4-point Likert scale to rate the favorable attitudes towards treatment in 10 items	University students (U) and trainee teachers (T)	M=25.20±3.34 (T) M=20.85±1.74 (U)	159 (T); 172 (U)	Singapore	A/H	Internal consistency; Structural validity
2. Han & Chan, 2015 [255] Attitudes Towards Help-Seeking Scale (ATHSS) Chinese version	4-point Likert scale to rate the favorable attitudes towards treatment in 29 items	Undergraduate students	M=19.64±1.42	353	China, Taiwan	A/H	Internal consistency; Reliability; Structural validity; Construct validity (known group validity)

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
3. Cakar & Savi, 2014 [254]	General Help Seeking Questionnaire (GHSQ)	7-point scale to ask respondents to rate the likelihood to seek help from a variety of people for 3 problem types	High school students	M=16.3±.70	198	Turkey	I	Internal consistency; Reliability
4. Cepeda-Benito & Short, 1998 [163]	Intention of Seeking Counseling Inventory (ISCI)	17 items on general problems students bring to counseling that are rated on 6-point Likert scale for likelihood to seek help	University students	M=19.5±1.98	732	US	I	Internal consistency; Structural validity (factor analysis); Construct validity
5. Deane et al., 2001 [164]	General Help Seeking Questionnaire (GHSQ)	7-point scale to ask respondents to rate the likelihood to seek help from a variety of people for 3 problem types	Undergraduate students	M=20.58±4.98	302	Australia	I	Internal consistency; Structural validity (factor analysis)
6. Elhai, et al., 2008 [165]	ATSPPH	4-point Likert scale to rate the favorable attitudes towards treatment in 10 items	College students (S); Patients (P)	M=20.7±3.3 (18-42)(S) M=47.3±17.7 (18-90) (P)	296 (S); 389 (P)	US	A/H	Content validity Internal consistency; Structural validity (factor analysis); Construct validity

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
7. Fischer & Turner, 1970 [162]	ATHSS	4-point Likert scale to indicate agreement on attitudes towards treatment in 29 items	High school, college and university students	N/P	N=492 female; N=468 male	US	A/H	Content validity; Internal consistency; Test-rest reliability; Structural validity (factor analysis); Construct validity
8. Fischer & Farina , 1995 [166]	ATSPPH	4-point Likert scale to rate the favorable attitudes towards treatment in 10 items	College students	Modal age=18	389	US	A/H	Internal consistency; Test-rest reliability; Structural validity (factor analysis); Construct validity; Criterion validity
9. Gulliver et al., 2012 [54]	ATSPPH	4-point Likert scale to rate the favorable attitudes towards treatment in 10 items	Elite athletes	M=25.5 (18-48)	59	Australia	A/H	Internal consistency; Test-rest reliability
	GHSQ	7-point scale to ask respondents to rate the likelihood to seek help from a variety of people for 3 problem types	Elite athletes	M=25.5 (18-48)	59	Australia	I	Internal consistency; Test-rest reliability

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
10. Hatchett et al., 2006 [167]	ATHSS	4-point Likert scale to indicate agreement on attitudes towards treatment in 29 items	University students	M=20.8 (18-53)	273	US	A/H	Internal consistency; Construct validity
11. Jorm, Blewitt et al., 2005 [168]	Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	One open ended question to examine how respondents assist a person described to have mental illness in vignettes	Community adults	≥18	3998	Australia	A/T	Inter-rater reliability
12. Jorm & Mackinnon et al., 2005 [169]	Jorm MHL	One open ended question to examine how respondents assist a person described to have mental illness in vignettes	Community adults	≥18	3998	Australia	A/T	Structural validity (factor analysis)
13. Komiya et al., 2000 [120]	ATSPPH	4-point Likert scale to rate the favorable attitudes towards treatment in 10 items	Undergraduate students	M=18.4±1.32	310	US	A/H	Internal consistency; Construct validity
14. Lee et al., 2014 [170]	Help Seeking Intentions (HSI)	3-point scale on 11 help seeking intentions	Children and youth	10-18	701	US	I	Internal consistency; Structural validity (factor analysis); Content validity

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
15. Mackenzie et al., 2004 [171]	The New Inventory of Attitudes Towards Seeking Mental Health Services (IASMHS)	4-point Likert scale to indicate agreement on attitudes towards treatment in 24 items	Undergraduate students	M=21±2.7	293	Canada	A/H	Content validity; Internal consistency; Test-retest reliability; Structural validity (factor analysis); Construct validity
16. Nickerson et al., 1994 [172]	Help-Seeking Attitude Scale (HSAS)	4-point Likert scale to rate agreement on help-seeking attitudes in 40 items	Black college students	M=20.29±2.6 17-37	105	US	A/H	Internal consistency; Concurrent validity
17. O'Connor & Casey, 2015 [219]	Mental Health Literacy Scale (MHLS)	Multiple choice on 35 items regarding knowledge and attitudes about help-seeking, and ability to recognize disorders	First year university students (S) Mental health professionals (M)	M=21.10±6.27 (S); M=33.09±8.01	372 (S); 43 (M)	Australia	A/K	Internal consistency; Reliability; Measurement error; Content validity; Structural validity; Construct validity

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
18. Reavley et al., 2014 [173]	Jorm Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	One open ended question to examine how respondents assist a person described to have mental illness in vignettes	Adults	≥15	7.555	Australia	A/T	Construct validity
19. Royal & Thompson, 2012 [174]	ATHSS	4-point Likert scale to indicate agreement on attitudes towards treatment in 29 items	Protestant Christians	>18	540	US	A/H	Structural validity (factor analysis)
20. Sahin & Uyar, 2011 [175]	Scale of Attitudes Toward Seeking Psychological Help for Secondary Students (ASPH-S)	Not reported (15 items)	High school students	14-19	301	Turkey	A/H	Internal consistency; Test-retest reliability; Structural validity (factor analysis)
21. Schmeelk-Cone et al., 2012 [176]	Help Seeking Acceptability (HSA)	4-point Likert scale to rate agreement on help-seeking acceptability in 4 questions	High school students	Not reported	6,370	US	A/H	Internal consistency; Structural validity (factor analysis); Construct validity; Content validity

Author/year	Measurement tool	Response options	population	Age	N	country	Content	Psychometric Properties Assessed
22. Tuliao & Velasquez, 2014 [256]	GHSQ-Philippine version	7-point scale to ask respondents to rate the likelihood to seek help from a variety of people for 3 problem types	Undergraduate students	M=17.69±.98	359	Philippine	I	Internal consistency; Reliability; Structural validity; Construct validity
23. Turner, 2012 [177]	Parental Attitudes Toward Psychological Services Inventory (PATPSI)	5-point Likert scale to rate agreement on help-seeking attitudes, intentions, and stigma	University students (S); Caregivers (C)	M=19.20±1.16 (S); M=34.41±6.1 (C)	250 (S); 260 (C)	US	A/H; I	Internal consistency; Test-retest validity; Structural validity (factor analysis); Construct validity
24. Wilson et al., 2005 [178]	GHSQ	7-point scale to ask respondents to rate the likelihood to seek help from a variety of people for 3 problem types	High school students (H); University students (U)	M=13.5; 15 (H) M=19 (U)	1,766	Australia	I	Internal consistency; Test-retest reliability; Construct validity; Content validity

A/H: beliefs/Attitudes towards help-seeking; A/K: attitudes towards help-seeking and knowledge where to seek help; A/T: beliefs/attitudes towards treatment; I: help-seeking intentions

Chapter 5 Table 2: Methodological Quality of a Study on Each Measurement Property of a Measurement Tool

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing (Construct validity)
1. Attitudes towards help-seeking scale (ATHSS)	Fischer & Turner, 1970 [162]	Fair	Fair	Excellent		Fair		Fair
ATHSS-Chinese	Han & Chan [253]	Good	Fair			Fair		Fair
ATHSS	Royal & Thompson, 2012 [174]					Good		
ATHSS	Hatchett et al., 2006 [167]	Poor						Fair
2. Attitudes Toward Seeking Professional Psychological Help Scale (ATSPPH)	Fischer & Farina, 1995 [166]	Good	Fair			Fair	Fair	Fair
ATSPPH	Elhai, et al., 2008 [165]	Good		Poor		Fair		Fair
ATSPPH	Ang et al., 2007 [253]	Good				Fair		
ATSPPH	Gulliver et al., 2012 [54]	Poor	Poor					
ATSPPH	Komiya et al., 2000 [120]	Poor						Fair
3. Intention of Seeking Counseling Inventory (ISCI)	Cepeda-Benito & Short, 1998 [163]	Good				Fair		Fair
4. General Help Seeking Questionnaire (GHSQ)	Deane et al., 2001 [164]	Good				Fair		

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing (Construct validity)
	Gulliver et al., 2012 [54]	Poor	Poor					
	Cakar & Savi, 2014 [254]	Poor	Fair					
	Tuliao & Velasquez [256]	Good	Fair			Fair		Fair
	Wilson et al., 2005 [178]	Poor	Fair	Poor				Fair
5. Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	Jorm, Blewitt et al., 2005 [168]		Good					
	Jorm & Mackinnon et al., 2005 [169]					Excellent		
	Reavley et al., 2014 [173]							Fair
6. Help Seeking Intentions (HSI)	Lee et al., 2014 [170]	Good		Poor		Fair		
7. The New Inventory of Attitudes Towards Seeking Mental Health Services (IASMHS)	Mackenzie et al., 2004 [171]	Good	Poor	Excellent		Good		Fair
8. Help-Seeking Attitude Scale (HSAS)	Nickerson et al., 1994 [172]	Poor						Fair
9. Mental Health Literacy Scale (MHLS)	O'Connor & Casey, 2015 [219]	Excellent	Good	Excellent	Good	Excellent		Fair

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing (Construct validity)
10. Scale of Attitudes Toward Seeking Psychological Help for Secondary Students (ASPH-S)	Sahin & Uyar, 2011 [175]	Good	Fair			Fair		
11. Help Seeking Acceptability (HSA)	Schmeelk-Cone et al., 2012 [176]	Good		Poor		Fair		Fair
12. Parental Attitudes Toward Psychological Services Inventory (PATPSI)	Turner, 2012 [177]	Good	Fair			Fair		Fair

Chapter 5 Table 3: Quality of Each Measurement Property

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
1. Attitudes towards help-seeking scale (ATHSS)	Fischer & Turner, 1970 [162]	+	+	+		?		+
	Han & Chan [255]	+	+			-		+
	Royal & Thompson, 2012 [174]					-		
	Hatchett et al., 2006 [167]	?						+
2. Attitudes Toward Seeking Professional Psychological Help Scale (ATSPPH)	Fischer & Farina , 1995 [166]	+	+			?	+	+
	Ang et al., 2007 [253]	+				?		
	Elhai, et al., 2008 [165]	+		?		?		+

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
	Gulliver et al., 2012 [54]	-	-					
	Komiya et al., 2000 [120]	?						+
3. Intention of Seeking Counseling Inventory (ISCI)	Cepeda-Benito & Short, 1998 [163]	+				+		+
4. General Help Seeking Questionnaire (GHSQ)	Deane et al., 2001 [164]	+				+		
	Gulliver et al., 2012 [54]	?	+/-					
	Cakar & Savi, 2014 [254]	?	+					
	Tuliao et al., 2014 [256]	+	+			+		+
	Wilson et al., 2005 [178]	?	+	?				+
5. Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	Jorm, Blewitt et al., 2005 [168]		+/-					
	Jorm & Mackinnon et al., 2005 [169]					-		
	Reavley et al., 2014 [173]							+
6. Help Seeking Intentions (HSI)	Lee et al., 2014 [170]	+		?		?		
7. The New Inventory of Attitudes Towards Seeking Mental Health Services (IASMHS)	Mackenzie et al., 2004 [171]	+	?	+		-		+
8. Help-Seeking Attitude Scale (HSAS)	Nickerson et al., 1994 [172]	?						+

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
9. Mental Health Literacy Scale (MHLS)	O'Connor & Casey, 2015 [219]	+	+	+	?	?		+
10. Scale of Attitudes Toward Seeking Psychological Help for Secondary Students (ASPH-S)	Sahin & Uyar, 2011 [175]	+	+			+		
11. Help Seeking Acceptability (HSA)	Schmeelk-Cone et al., 2012 [176]	+		?		+		+
12. Parental Attitudes Toward Psychological Services Inventory (PATPSI)	Turner, 2012 [177]	+	+/-			+		+

+: positive ratings; -: negative ratings; +/-: conflicting ratings; ?: indeterminate ratings

Chapter 5 Table 4: Overall Level of Evidence of Measurement Properties

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
1. Attitudes towards help-seeking scale (ATHSS)	Fischer & Turner, 1970 [162]; Han & Chan, 2015 [255] Hatchett et al., 2006 [167]; Royal & Thompson, 2012 [174]	+	+	+++		-		++
2. Attitudes Toward Seeking Professional Psychological Help Scale (ATSPPH)	Fischer & Farina , 1995 [166]; Elhai, et al., 2008 [165]; Ang et al., 2007 [253]; Gulliver et al., 2012 [54]; Komiya et al., 2000 [120]	++	x	x		x	+	++
3. **Intention of Seeking Counseling Inventory (ISCI)	Cepeda-Benito & Short, 1998 [163]	++				+		+

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
4. General Help Seeking Questionnaire (GHSQ)	Deane et al., 2001 [164]; Gulliver et al., 2012 [54]; Cakar & Savi, 2014 [254]; Tuliao & Velasquez, 2014 [256]; Wilson et al., 2005 [178]	x	+	x		++		++
5. Mental health literacy survey (items on attitudes/beliefs towards treatment) (Jorm MHL)	Jorm, Blewitt et al., 2005 [168]; Jorm & Mackinnon et al., 2005 [169]; Reavley et al., 2014 [173]		+/-			---		+
6. Help Seeking Intentions (HSI)	Lee et al., 2014 [170]	++		x		x		
7. The New Inventory of Attitudes Towards Seeking Mental Health Services (IASMHS)	Mackenzie et al., 2004 [171]	++	x	+++		--		+
8. Help-Seeking Attitude Scale (HSAS)	Nickerson et al., 1994 [172]	x						+
9. Mental Health Literacy Scale (MHLS)	O'Connor & Casey, 2015 [219]	+++	++	+++	x	x		+
10. **Scale of Attitudes Toward Seeking Psychological Help for Secondary Students (ASPH-S)	Sahin & Uyar, 2011 [175]	++	+			+		

Measurement tool	Study Author	Internal consistency	Reliability	Content validity	Measurement error	Structural validity	Criterion validity	Hypothesis testing
11. Help Seeking Acceptability (HSA)	Schmeelk-Cone et al., 2012 [176]	++		x		+		+
12. Parental Attitudes Toward Psychological Services Inventory (PATPSI)	Turner, 2012 [177]	++	+/-			+		+

+++=strong evidence with positive ratings; ----=strong evidence with negative ratings; ++ = medium evidence with positive ratings; -- = medium evidence with negative ratings; +=limited evidence of positive ratings; - = limited evidence of negative ratings; +/- = conflict findings; x = studies of poor methodological quality or studies with indeterminate ratings

**: measurement tools with all psychometric properties rated as limited evidence or above.

Appendix 2: Search Strategies in PubMed

Concept 1 AND **Concept 2** AND **Concept 3** AND **Concept 4**

	key Mental health disorders and mental health	**3 aspects of MHL**	**Assessment tool**	**Study type**
	"Mental Disorders"[Mesh: noexp] OR "mental	"health education"[tiab]	assessment*[tiab]	Reliability[tiab]
O R	health"[Mesh: noexp]			
	"Substance-related disorders"[Mesh] OR substance use disorder*[tiab] OR "substance abuse"[tiab] OR "substance misuse"[tiab] OR "substance dependence"[tiab]	"health education"[Mesh]	evaluat*[tiab]	effective*[tiab]
	anxiety disorder*[tiab] OR "anxiety disorders"[Mesh] OR "generalized anxiety disorder"[tiab] OR	"mental health literacy"[tiab]	measur*[tiab]	efficac*[tiab]
O R	"separation anxiety disorder"[tiab] OR			

	“social phobia”[tiab] OR “specific phobia”[tiab] OR “panic disorder”[tiab] OR “posttraumatic stress disorder”[tiab]			
	disruptive behavior disorder*[tiab] OR “attention deficit and disruptive behavior disorders”[Mesh] OR “conduct disorder”[tiab] OR “oppositional defiant disorder”[tiab]	“health knowledge”[tiab]	test*[tiab]	“program evaluation”[Mesh] OR “program evaluation”[tiab]
OR	“unipolar depression”[tiab] OR “major depressive disorder”[tiab] OR depression[tiab] OR “depressive disorder”[Mesh] OR “depression”[Mesh]	“health curriculum”[tiab]	scale*[tiab]	Validity[tiab]

OR	“attention deficit hyperactivity disorder”[tiab] OR ADHD[tiab]	“mental health awareness”[tiab]	assessment tool*[tiab]	
		awareness[Mesh]	psychometrics[Mesh] OR psychometrics[tiab]	
OR		“attitude to health”[Mesh]	questionnaires[Mesh] OR questionnaire*[tiab]	
OR			survey*[tiab]	
OR		stigma[tiab]		
OR		discrimination[tiab]		
		“help seeking behavior”[tiab] OR “seeking help”[tiab]		

Appendix 3: Quality Criteria of Measurement Properties

Property Quality	**criteria**	**Rating**
Reliability		
Internal consistency	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ $0.70	+
	Dimensionality not known OR Cronbach's alpha not determined	?
	(Sub)scale not unidimensional OR Cronbach's alpha(s) ,0.70	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Reliability	ICC/weighted Kappa ≥ $0.70 OR Pearson's r≥0.80	+
	Neither ICC/weighted Kappa, nor Pearson's r determined	?
	ICC/weighted Kappa ≤ 0.70 OR Pearson's r ≤ 0.80	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Measurement error	MIC > SDC OR MIC outside the LOA	+
	MIC not defined	?
	MIC ≤ SDC OR MIC equals or inside LOA	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Validity		
Content validity	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete	+
	No target population involvement	?

	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Structural validity	Factors should explain at least 50% of the variance	+
	Explained variance not mentioned	?
	Factors explain < 50% of the variance	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Hypothesis testing (construct validity)	Correlation with an instrument measuring the same construct ≥0.50 OR at least 75% of the results are in accordance with the hypotheses AND correlation with related constructs is higher than with unrelated constructs	+
	Solely correlations determined with unrelated constructs	?
	Correlation with an instrument measuring the same construct <0.50 OR <75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Criterion validity	Correlations with the gold standard is ≥0.70	+

	Correlations with the gold standard is unknown	?
	Correlations with the gold standard is <0.70	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-
Responsiveness		
Responsiveness	(Correlation with an instrument measuring the same construct ≥0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥0.70) AND correlation with related constructs is higher than with unrelated constructs	+
	Solely correlations determined with unrelated constructs	?
	Correlation with an instrument measuring the same construct <0.50 OR <75% of the results are in accordance with the hypotheses OR AUC <0.70 OR correlation with related constructs is lower than with unrelated constructs	-
	Positive rating (+) in one subgroup, however negative rating (-) or unknown (?) in another subgroup in the same study	+/-

Appendix 4: Level of Evidence for the Overall Quality of the Measurement Property

Level	Rating	Criteria
Strong	+++ or ---	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	+/-	Conflicting findings
Unknown	x	Studies of poor methodological quality or studies with indeterminate rating of the measurement property