Validation of the Recovery Attitudes Questionnaire and the Opening Minds Scale for Health Care Providers

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Psychology

at Massey University, Albany, New Zealand.

Esario IV Pueblo Daguman

2022

Acknowledgements

Taus-pusong pasasalamat kay Associate Professor Joanne Taylor, ang aking superbisor, para sa kanyang ekspertong paggabay, masiglang pagganyak, at masinsinang puna sa pananaliksik na ito.

Thank you to Associate Professor Joanne Taylor, my research supervisor, for her expert guidance, spirited motivation, and insightful feedback on this research.

Ang mga tao sa likod ng Postgraduate Research Fund at si Harvey Jones ay dapat ding kilalanin para sa kanilang tulong sa sample recruitment at instrumentation process.

The people behind Postgraduate Research Fund and Harvey Jones also need to be acknowledged for their assistance in sample recruitment and instrumentation.

Ang akademikong pagsusumikap na ito ay hindi magiging posible kung wala ang mga kalahok na nagbigay ng oras at suporta nila.

This academic endeavour would not have been possible without the participants who have given their time and support.

Para sa kanilang patuloy na suporta at pagpapalakas ng loob sa aking pag-aaral, nais kong magpasalamat sa aking mga magulang, sina Ruby at Esario Jr+, at mga kapatid na sina Esario III, Gladys, Gloiza and Zoe.

For their continued support and encouragement throughout my studies, I am deeply indebted to my parents, Ruby and Esario Jr+, and siblings, Esario III, Gladys, Gloiza and Zoe.

Panghuli, sa mga taong nasa kanilang paglalakbay tungo sa pagtulong sa kanilang sarili na maging mas mabuti at sa mga tumutulong sa kanila, ang aking paghanga at pagtitiwala ay nasa inyo. Salamat sa pagbibigay inspirasyon sa akin.

Lastly, to the people who are on their journey towards helping themselves become better, and to those who help them, my admiration and vote of confidence are with you. Thank you for inspiring me.

Abstract

A range of measures is used to evaluate interventions designed to shift stigmatising attitudes towards mental health recovery. However, there is mixed evidence on the psychometric properties of these measures and unvalidated measures continue to be used. Methodological issues may contribute to these mixed findings, such as the underreporting of critical measurement information. This study aimed to address these issues in evaluating the factorial validity and reliability of two measures of recovery and stigma attitudes: the 7-item Recovery Attitudes Questionnaire (RAQ-7; Borkin et al., 2000) and the 15-item Opening Minds Scale for Health Care Providers (OMS-HC-15; Kassam et al., 2012; Modgill et al., 2014), through a partial replication of Chiba et al. (2016) and Őri et al. (2020). The measures were completed online by 286 medical sector workers recruited from Prolific Academic, 19 of whom completed them again two weeks later. Confirmatory factor analyses and reliability estimations revealed that the RAQ-7 scores had an unsatisfactory internal and test-retest reliability, and poorly fitted the known two-factor structure. In contrast, the OMS-HC-15 scores demonstrated strong internal consistency, very weak test-retest reliability, mixed fit to the known correlated three-factor structure and weak to moderate support for the interrelationship between the factors. These findings indicate that a more valid and reliable alternative to the RAQ-7 must be used to measure recovery attitudes, whereas the OMS-HC-15 is a viable measure of stigmatising attitudes. Further robust and transparent psychometric validations are needed to integrate personal recovery and mental distress stigma measures into practice.

Preface

Mental disorders or mental illness is a vital issue historically and contemporarily. It is defined in many different ways (Telles-Correia et al., 2018) and is historically perceived as rare (Feldman & Crandal, 2007). However, much like physical illness, mental disorders are widespread globally and are experienced daily by people regardless of age, gender, ethnicity and socioeconomic background (e.g., Substance Abuse and Mental Health Services Administration, 2021; Vos et al., 2012). In New Zealand, one in every five adults is diagnosed as having problems with mood or anxiety (Ministry of Health, 2019).

Despite public health efforts, stigma and discrimination against people with mental distress persist (Rössler, 2016). Current thinking promotes the use of language that does not reinforce or contribute to stigma and discrimination (e.g., Gwarjanski & Parrott, 2018; Klin & Lemish, 2008; Volkow et al., 2021). There are words that best describe experiences of mental distress and the people who have them in a way that could change negative conceptualisations (e.g., Clement et al., 2010). Therefore, in the spirit of using less stigmatising words in messaging, from this juncture and throughout this Master's thesis, except where there is an exact quotation, the term "mental distress" is used instead of "mental illness" or "mental disorder". This decision has been made to reflect a broader understanding of mental distress. Furthermore, the words "people with lived experience of mental distress" or "people with mental distress" will be used instead of "people with a mental disorder" or "service users". It is understood that not all "people with lived experience of mental distress" may necessarily use services from the mental health system and be considered "service users". However, except for the brief discussion of the consumer movement in the first chapter, the words "people with lived experience of mental

distress" or "people with mental distress" are used to reflect the importance of focusing on the person at the front and centre of the experience of distress and the mental health services. Furthermore, such use signifies that the distressing experience(s) is only one part of that person's overall identity.

Time and again, different approaches to the conceptualisation of mental distress exist. Not everyone finds one way of thinking to be helpful. Similarly, people from diverse demographic backgrounds have varying ideas about mental health recovery and its barriers, such as mental distress stigma – two concepts that will be discussed and measured in this research. Therefore, the structure of this thesis will begin with the emergence and different conceptualisations of recovery and stigma relating to mental health. After discussing the two preceding concepts, the measures examined in this research will be introduced. Within the extended introduction, some of the issues in psychometric practice that have a significant implication for the accurate measurement of the two preceding concepts will be explored to guide the open and robust evaluation of the measures under study.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

What is Mental Health Recovery?

The language and meaning of mental health recovery have long been debated. A clinical and traditional approach, often defined by health care services and providers, describes recovery as a return to premorbid functioning, where symptoms are remitted or resolved (Kane, 2013; Liberman & Kopelowicz, 2005). In this perspective, recovery from mental distress is similar to recovery from physical illness, where diagnosis and treatment services are based on the medical model, with a focus on symptom reduction and remission (Jacob, 2015). This focus on services contributed to an increased stigma associated with mental distress and other impacts, such as reduced socioeconomic expectations and diminished hope for recovery (Harding et al., 1987). Consequently, individuals diagnosed with mental distress faced a dismal prognosis due to early assumptions of an inevitable downward trajectory and loss of overall functioning (Kruger, 2000). These attitudes permeated mental health care systems and the general public (Anthony et al., 1999).

While the medical model and clinical approaches to recovery predominate, there has been disagreement over the definition of recovery and the voices of consumers and service users have led the debate. People with lived experience of mental distress expressed their general dissatisfaction with and opposition to the clinical perspective and institutions that promulgated them, citing that it does not reflect lived experience values on empowerment and orientation to choice (Jacob, 2015). Those who were under the care of centralised structures of welfare, the service users or consumers, argued that they understood more about what they needed to

transform their lives (O'Hagan, 2009). While the clinical perspective remains necessary to some people with lived experience (e.g., some find value in diagnosis and treatment; Piat et al., 2009), others held a different perspective and approach to reclaim control and focus on the person with mental distress and not on the "illness" (Cohen et al., 2012). Most notably, the consumer movement was formed in the United States through the writings and activism of Chamberlin and other psychiatric 'survivors' during the late 1970s and grassroots organisations throughout the 1980s and early 1990s (Roberts & Wolfson, 2004). The 'survivors' wrote narratives on highly individualised outcomes that have challenged the prevailing belief that mental distress results in long-term and irreversible decline and has biological origins (e.g., Deegan, 1988; Leete, 1989; Unzicker, 1989). The Community Support System of the 1970s (Turner & TenHoor, 1978) and the ex-patient movement of the 1980s (Chamberlin, 1990) were organisations led by service users or consumers that helped define recovery by sharing their stories while supporting those who had lived experiences. Some psychiatrists and scholars rallied behind the movement and offered alternative views and strategies on how to cope with mental distress and its effects (Crossley, 2006). Furthermore, radical antipsychiatry emerged to actively oppose coercive and authoritative practices aimed at reducing mental distress rather than promoting hopes and dreams (Rissmiller & Rissmiller, 2006).

The consumer movement was vital for bringing the concept of personal recovery to the forefront of mental health care system changes (Cohen et al., 2012). Anthony (1993) defined recovery based on the accounts of people with lived experience, as follows:

> *Recovery is described as a deeply personal, unique process of changing one's attitudes, values, feelings, goals, skills, and/or roles. It is a way of living a satisfying, hopeful, and*

> *contributing life even with limitations caused by illness. Recovery involves the development of new meaning and purpose in one's life as one grows beyond the catastrophic effects of mental illness (p.15).*

This definition provided an important perspective on recovery, which has been used to inform mental health policy worldwide. In the 1990s, recovery-oriented practices were brought to the national policy level, with New Zealand as the first to mandate that services user or consumers should not only be allowed to participate in the pursuits of the mental health system but also lead and influence service development (New Zealand Mental Health Commission, 2001; O'Hagan et al., 2012; Tooth et al., 2003). The New Zealand Mental Health Commission (1998, p.113) defines recovery as "living well in the presence or absence of mental illness," echoing Anthony's 1993 publication that sparked US mental health reforms.

However, personal recovery is complex and multidimensional (Anthony, 1993; Mancini et al., 2005) and has been defined as a process (e.g., Jacobson, 2001), an outcome (e.g., Davidson, 2012), and an attitude (e.g., Anthony et al., 1999). People with lived experience define recovery differently, as do service providers, family, and communities (Pilgrim, 2008). Theories and models of personal recovery also take different approaches, and some but not all were developed by service users. Examples include but are not limited to the CHIME framework, which stands for connectedness, hope and optimism, identity, meaning in life, and empowerment (Leamy et al., 2011), the CHIME-D framework (where D stands for difficulties; Stuart et al., 2017) and the trauma-informed model (Ellison et al., 2018;). Other perspectives take issue with these models and definitions of recovery, pointing out that recovery is a very personal concept, reclaiming madness as an identity to be valued and challenging prevailing ways of being

"normal" and negative connotations of having lived experience of mental distress or being 'mad' (Beresford, 2020). In other words, for people with mental distress, there is a lack of consensus on what constitutes recovery, all the more so when one considers the definitions of those without lived experience of mental distress. However, in either of the perspectives briefly mentioned, Leonhardt et al. (2017) found two commonalities. Firstly, making sense of experiences of mental distress and making choices are important to recovery. Secondly, finding meaning can present several challenges that impact the likelihood of moving towards recovery. The individual will face struggles like the symptoms experienced and discouragement associated with mental distress.

Furthermore, a debate has occurred as to whether recovery is a process or an outcome (e.g., Bellack, 2006; Deegan, 1988; Spaniol & Koehler, 1994). The idea that recovery "from" mental distress or recovery as an outcome suggests that a person is returning to a healthy state if treatment milestones are achieved and symptoms are remitted (Davidson & Roe, 2007), consistent with the clinical perspective. In this view, clinicians are primarily responsible for determining whether a person is recovering (Kane, 2013). On the other hand, recovery "in" mental distress or recovery as a process means that recovery is an ongoing activity of meaning-making, where overcoming struggles and seeking fulfilment is the emphasis (Jacobson, 2001; Slade & Longden, 2015). It has been further argued that recovery as an outcome should not supersede the view of recovery as a process, considering that a person may have more to recover than just the effects of the distress (e.g., Law & Morrison, 2014; Slade & Longden, 2015). A person may still have to cope with the impact of the distress on the social or other dimensions of their life (Davidson & Roe, 2007). Recovery hence will have to be a continuous activity of understanding and overcoming pursuits more critical to the person (Piat & Sabetti, 2009). Thus,

a person's journey "in" living well does not stop when symptoms have remitted but often involves overcoming struggles and living a meaningful life despite enduring symptoms.

Many views have emerged since Anthony's (1993) writing on personal recovery. One is that recovery is a unique, nonlinear journey characterised by temporary setbacks and learning processes (Leamy et al., 2011). The importance of hope, the ability to be responsible for oneself, have meaningful activities, have a supportive social network, and have a positive identity also describe the recovery process and are considered the elements of recovery that form the CHIME conceptual framework (Leamy et al., 2011). Here, there is a striving to shift the focus from the distress (through the exercise of self-determination) to cultivating well-being (through the development of self-identity, valued roles and relationships; Slade, 2013). However, this view has been challenged as it is based on Western ideas of mental distress that value individuality and independence and omit consideration of collective values and spirituality (Slade et al., 2012). In other words, personal recovery as it is defined may not reflect cultural and minority groups who may view recovery collectively or differently.

Studies across the spectrum of mental distress describe recovery, as a unique process that may vary from person to person (Morrison et al., 2016). Some people with lived experience endorse the view of reclaiming control and distress (e.g., Spaniol et al., 2002), while others embrace the perspective of living well despite symptoms (e.g., Van Eck et al., 2018). Some may have no clear alternative to explain their "madness" and are specifically cautious of cancelling one overarching perspective to uphold another (Beresford, 2020). Recovery is also seen differently in specific contexts. Recovery is described holistically by people with schizophrenia (Jose et al., 2015). For people with personality disorders, emotional regulation improvement (Ng

et al., 2016), safety (Shepherd et al., 2016), and spirituality (De Ruysscher et al., 2017) have been some of the predominant themes. On the other hand, those recovering from eating disorders considered psychological wellness a key criterion for recovery (de Vos et al., 2017). In forensic populaces, a premium was given to specialised competence, trust, alternatives, and positive risk-taking (e.g., Clarke et al., 2016). The list could go on, but what do these all suggest? Recovery conceptualisations that people with lived experience of mental distress find acceptable are evolving and cut across any prevailing nomenclature (Andresen et al., 2003). A complex construct like recovery will be viewed differently and challenged continuously across different broad (i.e., transcending different mental distress) and specific (e.g., people with an eating disorder) contexts.

Measuring Personal Recovery

Various initiatives are being carried out to transform mental health services. These include the development of interventions (e.g., Bird et al., 2014; Chester et al., 2016) and policies both in New Zealand (Ministry of Health, 2012; O'Hagan et al., 2012) and globally (e.g., Commonwealth of Australia, 2009; Department of Health, 2001; Mental Health Commission of Canada, 2015) to ensure that the delivery of mental health services is oriented towards recovery that was borne out of the narratives of people with mental distress. At the core of some of these national initiatives (or mandates) is a set of domains, similar to earlier discussed elements of recovery (i.e., Leamy et al., 2011), that guide what form of help would be supportive to people with mental distress. What sets a recovery-oriented mental health service from other traditional practices is that it upholds hope, self-identity, self-determination, and personal responsibility, underpinned by an emphasis on connectedness (Slade, 2013). Consequently, recovery measures

will have to be developed and utilised to recognise that the focus of these mental health services is recovery (Davidson et al., 2007). However, despite rising recognition that recovery is a critical dimension of quality in mental health care, there has been a paucity of implementation and delivery of recovery-oriented mental health services in many countries, including New Zealand (Pincus et al., 2016). As an example, mental health service providers have difficulties reconciling what is required by the institution they work for and the needs of those they support (e.g., Le Boutillier et al., 2015). Furthermore, there is a need to improve the operationalisation of measures of service user-reported outcomes and recovery orientation of mental health care services, service providers or systems to fit with the policy mandates (Burgess et al., 2010; Stevens Manser et al., 2018).

Using appropriate recovery-oriented measures fulfils one of the many ways to establish a recovery-oriented system of care, as these measures aid in assessing the degree and progress of how well people, systems and organisations are in addressing recovery (White, 2007). However, choosing the appropriate recovery measure is not easy, given the complexity of recovery that has subsequently given rise to a range of measures of the different aspects of recovery (van Weeghel et al., 2019). Measures assess the effectiveness of interventions and services (as outcome measures), significant facets of the construct (as dimension measures) or the chronology of stages or a current location on the continuum of recovery (as process measure; Sklar et al., 2013). In addition, some recovery measures focus on either personal or organisational outcomes, where individual measures examine the service outcomes for people with mental distress, and organisational measures investigate the service user-centeredness of services (Stevens Manser et al. 2018). Given this range of measures, it is incumbent to thoroughly understand the validity and appropriateness of measures before employing them to evaluate recovery-oriented care and

services (Williams et al., 2012). Furthermore, selecting appropriate recovery measures may require that the measures are consistent with a person-centred approach, shifting the focus away from distress to being supported to make decisions about one's life (Davidson & Roe, 2007).

Another concern is that only a few measures of personal recovery have been developed and validated with the input of people with lived experience of mental distress. Most measures do not indicate how involved service users were in the development and evaluation process (Cavelti et al., 2012). If anything, only a few had people with lived experience of mental distress actively engaged throughout the process, and most definitions of recovery were variable and not first-hand (Sklar et al., 2013). Involving service users in the development and evaluation of such measures is critical because the recovery movement was primarily shaped by their experiences, which gained traction when their conceptualisations of personal recovery were highlighted. Integrating the experiences of people with mental distress (by including their definitions of recovery) into the recovery measures supports and affirms lived experience identity and meaning (Stuart et al., 2017). Moreover, placing lived experience voices at the centre of all aspects of mental health services or care would indicate a valued social role in maximising well-being (New Zealand Government, 2018).

In summary, recovery is a dynamic concept, and the experience of recovery remains unique to each individual. Given the diverse conceptualisations, people with lived experience of mental distress and mental health care service providers must collaboratively work on interpreting and assessing meaning-making to promote recovery (Leonhardt et al., 2017). Personal recovery needs to be measured to evaluate how policy mandates supporting recovery are implemented. However, a critical consideration in measuring personal recovery is ensuring

that the measure used is robust and maps onto recovery as a construct. Another important aspect in conceptualizing and measuring personal recovery is the concept and measurement of stigma, given that greater endorsement of recovery orientation is associated with less stigmatising attitudes (Stacy & Rosenheck, 2019).

What is Mental Distress Stigma?

A recovery approach takes a broader view than a symptom-focused one, emphasising what is meaningful for the person and supporting the development of valued social roles and relationships (Chester et al., 2016; Wood & Alsawy, 2018; Slade, 2013). Therefore, the approach is also consistent with reducing stigma regarding mental distress, which is held in society, services, and service providers, and internalised by people with lived experience.

The stigma associated with mental distress is a barrier to recovery for individuals with lived experience (Link & Phelan, 2001). People with mental distress can experience diminished social relationships or devalued identity if subjected to stigmatising encounters (Wahl, 2012). Due to stigma, people with mental distress can experience decreased confidence, optimism, self-reliance, esteem, and quality of life (Brohan et al., 2010; Corrigan et al., 2009; Livingston & Boyd, 2010). Stigma can exacerbate a host of socioeconomic and health burdens linked with mental distress (Hatzenbuehler et al., 2013). Individuals with mental distress are discriminated against in housing (e.g., Bambra & Pope, 2007; Corrigan et al., 2010; Link & Phelan, 2001), employment (e.g., Cechnicki et al., 2011; Sharac et al., 2010; Stuart, 2006), and medical areas (Rüsch et al., 2009; Thornicroft et al., 2007). Those who experience stigma have more severe symptoms (Boyd et al., 2014), seek fewer treatments (Corrigan, 2004), and do less well in therapy (Sirey et al., 2001).

Stigma has been defined in various ways (Bos et al., 2013). Goffman was one of the first to investigate stigma in 1963. In his view, stigma discredits and diminishes the value of a whole person to a minuscule part. It evolved into a language of relationships rather than an attribute, where one feature is not problematic (Jones, 1984). It also became associated with inequalities, cruelty, injustice, stereotypes, and subjective judgments (Crocker & Major, 1989). The stigma concept has been deconstructed into several subconstructs in psychology (Rao et al., 2019) and other disciplines. Recent years have brought a growing understanding of stigma as a multidimensional cultural phenomenon influencing every aspect of a person's life, including intrapersonal, interpersonal, organisational, and structural aspects (Abdullah & Brown, 2011; Hatzenbuehler et al., 2013; Yang et al., 2007).

Several models attempt to define and explain mental health stigma. According to Link and Phelan (2001), stigma occurs when *labelling*, *categorising*, *separating*, and *discriminating* combine within a power setting. Labelling starts with identifying key differences between the stigmatising person and the stigmatised person. Stereotyping then ensues when undesirable traits are associated with the earlier noted differences. These labelled individuals are subsequently categorised to achieve segregation and are allowed to lose status. As a result, stigmatised people can be discriminated against, stereotyped, and even eliminated. In this approach, the importance of attitudes and beliefs is thought to underlie the connections between the four components.

Corrigan's (2005) stigma conceptualisation has three parts. *Stereotypes* (about people with mental health issues), the first component, are mental representations based on prior experiences. This socially shared conceptualisation often presents people with mental distress with specific traits or functioning. *Prejudice*, the second component, is an unfavourable emotional reaction or attitude towards those with mental distress. *Discrimination*, the third

component, refers to the conduct that harms people with mental distress. As an example, a person who is recruiting for a role may think that people with mental distress are "incompetent" (stereotype), one of the most common stereotypes (Feldman & Crandall, 2007), may pity them (prejudice), the usual emotional response to the stereotype of competence (Corrigan, 2005) and deny them employment (discrimination), one of the many negative outcomes of stigma (Stuart, 2006).

Lastly, according to Thornicroft (2006), stigma comprises *information*, *attitudes*, and *behaviours*. In this instance, information refers to incorrect or insufficient knowledge based on culture or religion; attitudes refer to the harmful affective component, and behaviour refers to discrimination in Corrigan's (2005) model. This tri-partite model underlines the challenges that stigmatising attitudes bring and that almost everyone in society demonstrates these attitudes, including but not limited to health care service providers. Thornicroft (2006) noted that a lot of research on mental distress and stigma is conducted using measures of attitudes, which ask people about what they would do in an imaginary situation involving someone with mental distress. Items from these measures indicate that attitudes, which generate stereotypes and negative emotional responses, are linked to actual behaviours (Thornicroft & Kassam, 2008). The model assumes that attitudes, captured through a range of measures, are reflections of behavioural intents and could be used to measure several anti-stigma efforts.

In these three models, attitudes play an essential role in stigma. More specifically, attitudes towards people with mental distress can be located at the cognitive level and be quantified through stereotypes, consistent with other literature (e.g., Corrigan et al., 2002; Corrigan & Watson, 2002). It is also possible to see attitudes at the affective level and be

assessed through emotional responses and at the behavioural level as an intent, similar to the desire to maintain social distance toward people with mental distress (e.g., Corrigan et al., 2003).

Stigma can also be defined by its origin. Stigma regarding mental distress can originate from both the public and those who do not experience mental distress (Corrigan et al., 2005; Thornicroft et al., 2007), who can draw particular assumptions about someone based on their psychiatric diagnosis, such as people with mental distress are "dangerous", one of the core negative beliefs or stereotypes (Feldman & Crandall, 2007). Such beliefs that pervade society are *public stigma* (Michaels et al., 2017). This public stigma could be explained through three mechanisms: *stereotypes, prejudice*, and *discrimination*, based on evidence from social psychology (e.g., Brewer, 2007; Nelson, 2009) and stigma research (e.g., Bos et al., 2013; Pryor & Reeder, 2011). Stereotypes negatively deem a person with mental distress harmful, uncommon, unreliable, fragile, and reliant (Feldman & Crandall, 2007). On the other hand, prejudice can be manifested through expressions of anxiety (Hebl et al., 2000), fear, pity, and anger (Corrigan, 2005; Corrigan et al., 2004). Prejudice causes a person with mental distress to be denied opportunities or discriminated against through withholding help, avoiding, segregating and coercing (Corrigan & Rüsch, 2002; Corrigan & Watson, 2002).

Contrary to public stigma, *self-stigma* is characterised by accepting portrayals that dehumanise people with mental distress, including those that show the distress of others and oneself (Michaels et al., 2017). Self-stigma can be classified as *experienced*, *anticipated* and *internalised* (Fox et al., 2018). An experienced stigma comes with chronic injustices and acute, significant effects of stigma, which can have devastating consequences (Williams et al., 2003) as well as stereotypes, prejudices, and discrimination (e.g., Cechnicki et al., 2011; Quinn &

Earnshaw, 2011). Anticipated stigma occurs whenever someone with mental distress expects to be victimised by prejudice, discrimination, or stereotypes (Quinn & Earnshaw, 2011). Finally, internalised stigma includes people taking up the negative beliefs and feelings associated with the stigmatised identity (Bos et al., 2013; Corrigan et al., 2006; Link, 1987; Ritsher et al., 2003). However, it is essential to note that not all people with mental distress experience self-stigma and may validate, disagree with, or disregard the stigma that comes from outside them (Barney et al., 2010).

A specific case of stigma towards people with mental distress is held by mental health service providers. In their experience of seeking care, people with mental health distress regard mental health service providers as a substantial source of stigma (Schulze, 2007). While service providers may synthesise knowledge and representations about a mental health condition from various sources (Morant, 2006), they are likely to hold the same stigmatising views of mental distress as the public (e.g., Chester et al., 2016). They may have low expectations for both recovery and the consumer as a person (e.g., Berry et al., 2010). These biases can hinder the empowerment of people with mental distress and impact their recovery (Roberts & Wolfson, 2004). Negative attitudes of care providers may lead them to consider psychiatric illnesses and consumers as critical barriers to treatment (Chester et al., 2016; Salyers et al., 2009). Negative attitudes among care provider leadership may also alter the focus and pace of recovery initiatives (Le Boutillier et al., 2015; Piat & Lal, 2012). Therefore, anti-stigma programmes target healthcare workers and mental health professionals (Arboleda-Flórez & Stuart, 2012; Corrigan, Michaels et al., 2012).

Measuring Mental Distress Stigma

Several policy interventions have been undertaken to protect and normalise stigmatised groups in New Zealand (Dalziel, 2001) and other countries around the world (e.g., Cook et al., 2014; Pietrus, 2013). A common goal of these policies is to deliver measurable improvements in the social inclusion of people with mental distress. These interventions take the form of education and contact with people with mental distress and can be regarded as efforts to promote more inclusive attitudes in the general public, including mental health care professionals (e.g., Borschmann et al., 2014; Griffiths et al., 2014; Livingston et al., 2014). If combined, not only reductions in stigma but also improvement in recovery attitudes were found when health professionals were in contact with people with lived experiences of mental distress who shared their recovery stories (e.g., Taylor & Gordon, 2020), consistent with anti-discrimination contact-based education with a focus on recovery (Corrigan, Morris et al., 2012; Stuart et al., 2014). The key ingredient common in most of these contact-based education programmes targeted at health care professionals includes a trained speaker with lived experience of mental distress sharing their journey towards recovery (Knaak et al., 2014). Despite the widespread use of anti-stigma interventions, people with lived experience of mental distress, their advocates and stigma academics argue that changes should not focus on shifting attitudes alone but on legislative changes to improve the overall quality of life and social equality for the stigmatised (Stuart et al., 2012; Thornicroft et al., 2007). Notwithstanding what the stigmatised and the expert population view as more important, anti-stigma interventions often measure attitudes towards mental distress, so the measurement of stigma is an important area of research.

A particular measurement issue is the inherent complexity and multidimensionality of the construct of mental distress stigma (Brohan et al., 2010; Livingston & Boyd, 2010). Various

disciplines have developed definitions of stigma, terms representing an aspect of it, subsequent interventions that specifically address it, and measures to evaluate the intervention (Bos et al., 2013). The growing number of measures seem to have stemmed from the appeal to simplify the multifaceted construct (Fox et al., 2018). A considerable proportion of these measures have focused on evaluating stereotypes, the attitude component (Thornicroft & Kassam, 2008), given the need to address attitude change to reduce stigma. It is argued that this approach is practical because it requires focusing on specific behaviour and noticing the complexities of the beliefs and attitudes of an individual, which leads them to prejudice and discrimination (Link & Phelan, 2001). However, the overabundance of specific measures of stigma could be problematic in facilitating the selection of measures appropriate for an assessment need. Having several measures for an aspect of the same construct when existing measures are already available may be inefficient, as it hinders researchers' ability to draw general conclusions from the literature (Corrigan, Michaels et al., 2012). In addition, targeting and measuring one aspect of the complex stigma construct could lead to overlooking some of the other mechanisms in the larger stigma construct that may reinforce the target aspect of the construct (Link & Phelan, 2001). For example, if the aspect of "incompetence" in employment is used, one of the core stereotypes associated with people with mental distress (Feldman & Crandall, 2007), then the likely outcome would be excluding that "incompetent" person from being considered for a job, a common negative outcome traced to recruiters' stereotypes (Rooth, 2010). This kind of approach might help in understanding some aspects of mental distress stigma, but it could also prevent one from seeing the full consequences of it, such as the feelings of pity that the recruiter may have, a common negative feeling associated with the stereotype of "incompetence" (Corrigan, 2005). In other words, this approach of targeting and measuring specific aspects of the stigma construct

means that there are a lot of hypothetical situations involved and that the full context of the stigmatising experience is not captured (i.e., including emotions), which subsequently does not provide practical implications on how to reduce discriminating outcomes (Thornicroft, 2006). Therefore, to effectively address mental distress stigma, the different mechanisms and factors that engender and perpetuate it and its outcomes need to be understood and measured (Link et al., 2004).

Summary

There have been legislative measures to address identified needs and changes in mental health care systems, such as shifting stigmatising attitudes toward personal recovery. Consequently, it is important to use valid measures of the personal recovery and mental distress stigma constructs to work toward these recovery-oriented, anti-stigma goals. However, evaluating these constructs is challenging because they are complex and multidimensional. In addition, these constructs are seen differently by people with lived experience of mental distress and those without, particularly mental health service providers. Furthermore, despite the range of measures available for evaluating an aspect of these complex constructs and other specific purposes, these measures are not necessarily well-validated or comprehensive. The present study contributed to this literature by validating two such measures of attitudes to recovery and stigma, respectively– the Recovery Attitudes Questionnaire (RAQ; Borkin et al., 2000) and the Opening Minds Scale for Health Care Providers (OMS-HC; Kassam et al., 2012; Modgill et al., 2014).

The Recovery Attitudes Questionnaire and the Opening Minds Scale for Health Care Providers

Two measures of attitudes to personal recovery and mental distress stigma, respectively, are the RAQ and the OMS-HC. Many researchers have employed these measures conjointly to assess the efficacy of interventions to reduce negative attitudes to recovery and stigma across different groups and disciplines of health service providers (e.g., Foster et al., 2019; Mötteli et al., 2019; Sutton & French, 2019). In New Zealand, these two measures were utilised to evaluate the effectiveness of service user-led teaching that promoted mental health concepts to medical students (Gordon et al., 2014; Gordon et al., 2021; Newton-Howes et al., 2021) and postgraduate psychology students (Taylor & Gordon, 2020). The pre-test scores reflected attitudes before any interventions, whereas the post-test scores were used to detect change over time, mainly after user-led teaching. While studies support the psychometric properties of these two measures, there are inconsistencies regarding their internal structure, internal consistency, and temporal reliability.

Recovery Attitudes Questionnaire

Overview. Borkin et al. (2000) designed the RAQ to assess attitudes toward recovery from mental distress. The authors anchored the central construct on the recovery model of Anthony (1993), which emphasises the non-linearity of recovery, its influence on symptomatology, and its independence from professional intervention. In the US, a team of mental health consumers, professionals, and graduate students at Hamilton, Ohio, designed the RAQ to determine whether mental health service users would benefit from empowerment-based interventions over traditional interventions, such as therapy. It is a recovery outcome metric

(Stevens Manser et al., 2018) that examines attitudes towards rather than specific categories of personal recovery (Burgess et al., 2010). Self-report items include "People in recovery sometimes have setbacks" and "People with serious mental illnesses can strive for recovery." Responses are made on a five-point Likert scale from strongly disagree (1) to strongly agree (5). The RAQ does not have norms, but higher scores suggest a more positive attitude toward recovery.

The RAQ is copyright-free (Stevens Manser et al., 2018). There is no official placeholder for the measure or its manual (e.g., website). However, the items in the measure are part of the published work of the authors. Because the authors intended to create a user-friendly measure, it does not require training to use or administer. There is no indication of the required reading level. The RAQ should take five minutes to be completed (Gyamfi et al., 2021).

Development. Mental health service users in the US were involved in the development of the RAQ, from item analysis through to validation. Items were collected from their recovery writings and personal accounts and reviewed by other people with lived experiences of mental distress. The preliminary item pool consisted of 21 items. These items were tested with service users' families, mental health professionals, students and members of the general public. The RAQ was administered in various settings, including mental health forums, agencies, schools, and public venues, such as malls in small towns.

Borkin et al. (2000) conducted several item and factor analyses with groups of different ages, genders, races, mental health histories and lengths of time working with people with lived experience of mental distress. Factor loadings between .4 and .6 were maintained (Borkin et al., 2000). Analyses revealed that respondent groups viewed recovery and RAQ items differently. As a result of item overlap, the 21 items were reduced to 16 and distributed over four scales.

However, the iterative analytical process was initially expected to streamline the measure to reflect the dimensionality of recovery attitudes and be more reliable and easier to complete. Thus, the authors further shortened the measure to seven items which load to a two-factor structure - 1) "Recovery is possible and needs faith" and 2) "Recovery is difficult and differs among people" – that explained 54% of the variance in data (Borkin et al., 2000).

Validation. Table 1 summarises the research on the factorial validity, internal consistency and test-retest reliability of the RAQ-7 across different countries with various samples. It includes the results of studies that evaluated a range of psychometric properties of the measure and studies that used the measure to evaluate intervention and reported a single property, such as internal consistency. The discussion of the psychometric properties of the RAQ-7 follows the order of how they are presented in Table 1 (i.e., from left to right; from validity evidence to reliability estimates).

There have been limited factorial validity analyses after the validation of Borkin et al. (2000). Only one investigation found that the known two-factor structure for the RAQ-7 fitted the data well (Chiba et al., 2016). However, three items (2, 5, and 7) exhibited low factor loadings, and the incremental fit index did not reach the cutoff value in the same study. Other studies failed to support the known structure, even stating that the RAQ-7 is unidimensional (Hungerford et al., 2015; Wilrycx et al., 2012). The unidimensionality was attributed to several reasons, including the translation, sample homogeneity, and absence of service user-centred orientation in mental health care (Wilrycx et al., 2012). Further, in all studies that investigated factorial validity, item 2, "To recover requires faith", showed a low factor loading, raising the question of whether such a measure should be employed in populations outside of America (Chiba et al., 2016; Hungerford et al., 2015; Wilrycx et al., 2012). The disagreement in the

construction of "faith" was attributed to the absent or vague definition of constructs in the initial development and evaluation study for the RAQ-7, like that of "attitudes" (Hungerford et al., 2015).

The reliability of the RAQ-7 has predominantly been demonstrated in terms of internal consistency and has been found to have mixed internal reliability estimates, ranging from undesirable to satisfactory (α = .57 - .74). In contrast, test-retest reliability has been uniformly depicted as less than good (r, ICC and κ values less than .7). It was predicted by Borkin et al. (2000) that the RAQ-7 would generate higher internal consistency estimates when validated with homogeneous samples. However, Wilrycx et al. (2012), whose samples are less heterogeneous than the initial validation (i.e., mental health professionals and consumers only) and did not estimate retest reliability, suggested that retest reliability rather than internal consistency should be optimised in the RAQ-7.

Issues. There is no consensus on the findings among the RAQ-7 validation studies and the measurement practices employed therein, as seen in the different metrics and figures presented in Table 1. While there are no uniform standards on the latter due to the researcher's freedom and diversity of research problems, the validation studies were sparse in number and reasoning (on their chosen metrics, cutoffs, models and decision-making procedure). These disparities in practices occur frequently and are problematic. First, there seems to be a difference in the appreciation of test-retest reliability as a more important index of reliability over internal consistency among the validators of the RAQ-7, despite the mixed findings on the latter that are all reported using Cronbach's α (see Internal Consistency column in Table 1). This point reflects the over-reliance on internal consistency as the only estimate of reliability and Cronbach's α as the only estimate of internal consistency, supporting earlier findings that reflect the common yet

not rigorous way of estimating reliability (e.g., McCrae et al., 2011; Rammstedt & Beierlein, 2014; Sijtsma, 2009). Such an approach often results from selecting the incorrect measurement precision barometer (e.g., internal consistency) and the failure to understand that such a barometer need not necessarily be high if the construct to be measured is expected to change over time after intervening efforts (Kruyen et al., 2012). Also, using Cronbach's α is problematic as it assumes that there is only one factor within a scale and all the items' factor loadings are the same or tau equivalent (Schmitt, 1996), which is not the case for measures like the RAQ. Second, the methodological difference between Wilrycx et al. (2012) and Chiba et al. (2016) raises questions about whether exploratory and confirmatory processes should be done in the same data to prevent overfitting hazards. The study with both processes performed in the same dataset bears the danger of comparing data too closely, as it cannot analyse new data or forecast future observations (e.g., Fokkema & Greiff, 2017; Knekta et al., 2019; Ziegler, 2014). Third, the validation studies used different model fit indices, complicating information synthesis for entry-level researchers or readers of the general public. Such a complicated synthesis process would entail that more time and resources are needed, given that many researchers do not have the necessary background in psychometrics to make an unbiased assessment of the structural validity of a measure (Hussey & Hughes, 2020). Fourth, Wilrycx et al. (2012) did not provide their inference criteria; therefore, it is unclear how they decided whether their model misspecification hypotheses were validated. Although it is allowed for researchers to apply their methods when it comes to making decisions regarding the use of measures in a study (Vandenberg & Lance, 2000), unclear guidelines could lead to multiple interpretations of a given finding. Fifth, Chiba et al. (2016) concluded that the two-factor structure fitted their data based only on the absolute fit indices, confusing research users with their other equivocal findings. One fit index does not

cover the other equivalently important aspects of model-data fit, including fits sensitive to model size and non-normality (DiStefano & Hess, 2005). Finally, Hungerford et al. (2015) did not explain how they concluded that their "data did not fit well within the two factors" (p. 173), leaving readers to wonder if the model was ever fitted to their data. Such underreporting of evidence of the structural validity of a measure brings into question whether there could be some other psychometric properties of the measure that are invalid and hiding in plain sight (Flake et al., 2017).

Table 1

Summary of the Psychometric Properties of the RAQ-7

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
Borkin et al., 2000 (United States)	Service users, their families, mental health professionals from various fields, students, and members of the general population	844	PCA: 2 factors	Total: α = .70, Factor 1: α = .66, Factor 2: α = .64	Total: r = .67, Factor 1: r = .61, Factor 2: r = .62 (over 19 days)
Hardiman & Hodges, 2008 (United States)	Service providers from social work, psychology, and psychiatry	301		Total: α = .72	
Salgado et al., 2010 (Australia)	Service providers from government and non-government organisations	103		Total: α = .72	
Wilrycx et al., 2012 (The Netherlands)	Mental health professionals from various disciplines and settings and mental health consumers	210	CFA: 2 factors – SRMR = .064, RMSEA = .119, TLI = .645 EFA: 1 factor - Heywood case (i.e., only item 6 loaded to Factor 2)	Total: α = .61	

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
Jaeger et al., 2013 (Switzerland)	Acute psychiatric ward patients	81		Total: α = .57, Factor 1: α = .62, Factor 2: α = .28	
Rabenschlag et al., 2014 (Switzerland)	Inpatients and outpatients with various psychiatric diagnoses	149		Total: α = .68, Factor 1: α = .73, Factor 2: α = .77	
Hungerford et al., 2015 (Australia)	Service providers from publicly-funded bed-based and community-located mental health service	58	No explicit mention of the factor analysis method was made. "Data did not fit well within the two factors (Hungerford et al., 2015, p. 173)."	Total: α = .74, Factor 1: α = .72, Factor 2: α = .67	
Chiba et al., 2016 (Japan)	Mental health service providers from psychiatric hospitals, clinics, and community service agencies	307	SEM: 1 factor – GFI = .93, AGFI = .85, CFI = .77, AIC = 111.85 Two factors – GFI = .95, AGFI = .90,	Total: α = .64, Factor 1: α = .53, Factor 2: α = .56	Total: ICC = .68, Items 3 & 7: κ = .41-.6, Items 1, 2, 4, 5 & 6: $\kappa \leq .4$ (over 2 weeks)

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
			CFI = .86, AIC = 85.26		

Note. α = Cronbach's alpha, AGFI = Adjusted Goodness of Fit Index, AIC = Akaike information criterion, CFA = Confirmatory Factor Analysis, CFI = Comparative Fit Index, EFA = Exploratory Factor Analysis, GFI = Goodness of Fit Index, ICC = Intraclass Correlation Coefficient, κ = Cohen's kappa, PCA = Principal Component Analysis, r = Pearson correlation coefficient, RMSEA = Root Mean Square Error of Approximation, SEM = Structural Equation Modelling, SRMR = Standardised Root Mean Squared Residual, TLI = Tucker-Lewis Index. The data presented in the table are the aggregate data rather than data for each group in that study.

Opening Minds Scale for Health Care Providers

Overview. The OMS-HC was developed as a measure of stigmatising attitudes by Kassam et al. (2012) as part of the anti-stigma initiative "Opening Minds" of the Mental Health Commission of Canada. This measure was established to assess the impact of Opening Minds' contact-based educational programmes, where people with mental distress share their stories with target audiences. The OMS-HC developers derived their definitions from several models of mental health stigma, including Link and Phelan (2001), Corrigan (2005), and Thornicroft (2006). This self-administered test contains items like "I would be reluctant to seek help if I had a mental illness" and "There is little I can do to help people with mental illness." Response options range from strongly disagree (1) to strongly agree (5) on a 5-point Likert scale. The OMS-HC has no normative data. Lower scores suggest less stigmatising attitudes. The measure is available in the public domain. However, there is no online manual for administration, scoring and interpretation.

Development. The OMS-HC items were developed through several processes, including adapting existing scales' items, consultation with stigma intervention professionals, interviews with volunteers, and focus group discussions with diverse participants of different ages and genders and health care professional groups. To create an item pool, the authors developed new items and adapted some from the Medical Conditions Regard Scale (Christison et al., 2002), a measure used to assess the attitudes and emotional responses of physicians toward their patients. After generating the items, interviews were conducted to evaluate the sources of response error in the measure. The developers also conducted focus groups to discuss various topics, such as the definition of "mental illness", "recovery", and "social desirability". They further discussed whether or not the measure should include a definition of these terms and whether it should be

specific for a certain severity of mental distress. Although there was no consensus on these topics, the discussion was still valuable in refining the items. Throughout the development phase, the authors considered it crucial to involve people with lived experience of mental distress to ensure that concepts vital to them would not be missed out and that the wording of the measure was appropriate. The development phase led to creating a refined measure with 20 items.

The measure was initially tested with samples from different academic and professional organisations throughout various regions in Canada. Social desirability testing, item analysis and iterative factor analyses were performed. Six items exhibited poor item-total correlation values in the item analysis and were removed. After removing these items, factor analysis showed a three-factor structure. However, the third factor had only two items loaded onto it, and these two were then eliminated, reducing the OMS-HC to 12 items. The items loaded to two factors - 1) attitudes of healthcare providers towards people with mental illness and 2) attitudes of healthcare providers towards disclosure of mental illness.

The earlier iterative factor analysis ruled out a unidimensional structure and the third important factor of social distance. The developers of the OMS-HC wondered why this factor did not materialise despite inputs from the earlier consultation with experts in mental health stigma. Experts noted that social distance is a proxy for the behavioural intention to discriminate against people with lived experience of mental distress and the unwillingness to engage with them in different social relationships and activities. The developers regrouped to revisit the structure of the measure with a more representative respondent pool. A 15-item version (Modgill et al., 2014) was developed with a more stable three-factor structure. The three factors were (re)named *Attitude, Disclosure and Help-seeking,* and *Social Distance.*

Validation. Table 2 summarises the factorial validity, internal consistency and test-retest reliability of both the English and translated versions of the OMS-HC-15. Similar to the RAQ-7, the discussion of the psychometric properties of the OMS-HC follows the same order as to how they are presented in Table 2 (i.e., from left to right; from validity evidence to reliability estimates).

Numerous validators have claimed that their data have confirmed the known three-factor structure of Modgill et al. (2014). However, not all have established explicit inference criteria for determining whether a subset of fit indices must be met to confirm the model. Researchers supported the structure differently, based on all fit indices meeting the cutoff values (Van der Maas et al., 2018) and one absolute fit index and one relative fit index only (Sapag et al., 2019; Zuaboni et al., 2021). Other validators deleted 1 (Chang et al., 2017) to 2 (Happell et al., 2019) items but found similar equivocal outcomes. Another research group removed an item (i.e., Item 14, “More than half of people with mental illness don't try hard enough to get better") from the OMS-HC because it did not load well to the known structure (Őri et al., 2020; Hungarian study). The same Hungarian group then proposed a bifactor solution with items loading to both a general factor (i.e., the *overall stigmatising attitude*; p.5) and the specific factors (i.e., the known three factors of Modgill et al., 2014). Unfortunately, this alternative structure lacked model-based reliability (Őri et al., 2020).

The overall OMS-HC internal consistency estimates were strong (α = .69 - .79), while its subscale scores ranged from moderate to very strong (α = .48 - .83). The factor 2 subscale, *Disclosure and Help-seeking*, which has the fewest items, has been shown to have the lowest internal consistency of the subscales, also ranging from moderate to very strong (α = .48 - .83).

On the other hand, although only examined twice, the test-retest reliability of the OMS-HC-15's total (r = .99; ICC = .95) and subscale scores (ICC = .84 - .9) was very strong.

A noteworthy discovery outside of the validation studies' factor analytic work and reliability estimation was the emergence of a "shared conceptual theme" (or more like a nomological network[1]; Modgill et al., 2014, p. 8). The developers of the OMS-HC-15 discovered that the relationship between the known three factors in the measure could be linked to the convergent associations between stigmatising attitudes, social distance and help-seeking found during the development and validation of the Self-Stigma of Depression Scale (Barney et al., 2010), a measure of self-stigmatising attitudes about depression. Barney et al. (2010) revealed that when people view themselves as being unworthy and inferior (consistent with some items in the OMS-HC *Attitude* subscale), they tend to fear seeking help (similar to some items in the OMS-HC *Attitude* subscale). Subsequently, they noted that when people fear seeking help, they anticipate being viewed negatively by others (reflective of the factor measured by the OMS-HC *Disclosure and Help-seeking* subscale) and avoid interacting with them (representative of the factor measured by the OMS-HC *Social Distance* subscale). In addition, the foregoing convergent associations also aligned with the view of Corrigan and Watson (2002) that self-stigma may result from the acceptance of public stigma. Őri et al. (2020, p.7) demonstrated support for the notion of a "shared conceptual theme" (Modgill et al., 2014, p. 8) when they found significant (although weak) correlations between the three factors for the OMS-HC. They

[1] A nomological network is a depiction of a study's constructs, their observable manifestations, and their interrelationships (Preckel & Brunner, 2017).

concluded further that mental health providers are similar to everyone else who could experience distress, potentially isolate themselves and be less likely to ask for help.

Issues. Compared to the RAQ-7, the OMS-HC-15 has been more frequently studied. Following Modgill et al. (2014), numerous validations of the translated versions were diverse in findings and methodology, which is evident in the number of psychometrics studies and the different metrics used to present a reliability or validity estimate shown in Table 2. For instance, factor analysis was done separately for each professional group of health care providers in Modgill et al. (2014) but as an entire heterogeneous group in the subsequent studies. Such exclusion of tests that determines if the measurement model is equivalent across different groups could not help establish whether the measured construct could hold similar meaning for the populations (DiStefano & Hess, 2005). In other words, there is not sufficient information to help qualify that the known three-factor model for the OMS-HC true to, say, a sample of physicians and psychologists could also be valid to a sample of physicians only because previous factor analyses were mostly not stratified to groups. Second, the validations used different metrics (i.e., absolute and incremental fit indices) in their confirmatory analyses but had not discussed how they arrived at these methodological choices. While multiple fit indices provide a wealth of information regarding other equivalently important aspects of model-data fit (Russell, 2002; Schumacker & Lomax, 1996), the absence of information on whether fit indices come from the same family (e.g., if they all measure absolute fit only or incremental fit only) could not help those who are not experts in psychometrics determine that these are comprehensive. For example, a researcher without a strong foundation in psychometrics would not understand if the findings from, say, Sapag et al. (2019) would be more comprehensive in showing validity evidence of the OMS-HC than Őri et al. (2020), or vice-versa. Lastly, while the validations

mainly used the same set of model fit cutoffs (i.e., Hu & Bentler, 1999), the interpretations of these guidelines varied and were unexplained. Hu and Bentler (1999) endorsed employing two fit indices rather than one to decrease Type I and Type II errors. They suggested that model fit be evaluated using standardised root mean squared residual (SRMR) plus root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) meeting cutoff values. Such recommended appropriate use, combination and reporting of indices to support model fit was not applied in Őri et al. (2020), as they did not employ SRMR, and in Sapag et al. (2019), as they conclude that the OMS-HC was structurally valid based on absolute fit only albeit using the same set of fit indices from Hu and Bentler (1999).

Table 2

Summary of Psychometric Properties of the OMS-HC-15

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
Modgill et al., 2014 (Canada)	Health care providers	1,523	EFA: 3 factors, with 15 items	Total: α=.79, Factor 1: α = .68, Factor 2: α = .67, Factor 3: α = .68	
Chang et al., 2017 (Singapore)	Health care students (i.e., medicine & nursing)	1,002	CFA: 3 factors, with 15 items – RMSEA = .096, CFI = .838, TLI = .804 EFA, ESEM: 3 factors, with 14 items (less item 1) – RMSEA = .069, CFI = .948, TLI = .909	14 items (less item 1) – Total: α = .75, Factor 1: α = .74, Factor 2: α = .60, Factor 3: α = .53	
Destrebecq et al., 2018 (Italy)	Health care students (i.e., nursing, physical therapy, dietetics, occupational therapy)	561	EFA: 3 factors, with 15 items	Factor 1: α = .76, Factor 2: α = .83, Factor 3: α = .82	Total: r = .99 (over one week)

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
van der Maas et al., 2018 (Canada)	Community health centre staff	190	CFA: 3 factors, with 15 items – SRMR = .07, RMSEA = .013, CFI = .996, TLI = .996	Total: α = .77, Factor 1: α = .79, Factor 2: α = .67, Factor 3: α = .72	
Happell et al., 2019 (Norway, Australia, Finland, Ireland, and the Netherlands)	Nursing undergraduate students	423	CFA: 1 factor, with 15 items – RMSEA = .11, WRMR = 1.7, CFI = .79, TLI = .75 CFA: 3 factors, with 15 items – RMSEA = .07, WRMR = 1.17, CFI = .91, TLI = .89 CFA: 3 factors, With 13 items (no item 6 & 18) – RMSEA = .06, WRMR = 0.98, CFI = .95, TLI = .93	13 items (less item 6 & 18) – Factor 1: PSI = .61, Factor 2: PSI = .50, Factor 3: PSI = .69	

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
			Rasch analysis: 3 factors, with 13 items (no item 6 & 18) – Factor 1: $\chi 2(30) = 54.47$, $p = 0.004$, Item fit residual *SD* = 1.21, Person fit residual *SD* = 1.05 Factor 2: $\chi 2(12) = 12.42$, $p = 0.412$, Item fit residual *SD* = 1.62, Person fit residual *SD* = 1.25 Factor 3: $\chi^2(30) = 41.06$, $p = 0.086$, Item fit residual *SD* = 0.52, Person fit residual *SD* = 1.05, Unidimensionality test = 4.02%		
Sapag et al., 2019 (Chile)	Public health care staff and providers	803	SEM: 3 factors, 15 items – SRMR = .048, RMSEA = .052, CFI = .832, TLI = .798	Total: $\alpha = .69$, Factor 1: $\alpha = .53$, Factor 2: $\alpha = .48$, Factor 3: $\alpha = .60$	

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
Őri et al., 2020 (Hungary)	Professionals in psychiatry	211	CFA: original 15 items – RMSEA = .048, CFI = .818, TLI = .78 CFA: 1 factor, with 15 items – RMSEA = .066, CFI = .642, TLI = .583 CFA: 3 factors based on EFA, with 15 items – RMSEA = .045, CFI = .844, TLI = .812 CFA: 3 factors based on EFA, with 14 items (less item 14) – RMSEA = .043, CFI = .867, TLI = .836 CFA: bifactor model (1 general & 3 specific factors),	Model-based reliability of bifactor model, with 14 items (less item 14) – Total: ECV = .43, ω_h = .56; Factor 1: ECV = .18, ω_h = .37, Factor 2: ECV = .19, ω_h = .44, Factor 3: ECV = .19, ω_h = .37, PUC = .71 Simulation with α, with 14 items (less item 14) – Total: α = .73, Factor 1: α = .54, Factor 2: α = .63, Factor 3: α = .66 Correlations between factors: General to Factor 1: r_s = .68 General to Factor 2: r_s = .69 General to Factor 2: r_s = .73 Factor 1 to Factor 2: r_s = .22 Factor 1 to Factor 3: r_s = .33	14 items (less item 14) – Total: ICC = .95, Factor 1: ICC = .9, Factor 2: ICC = .88, Factor 3: ICC = .84 (over 1 month)

Researchers, Year of Publication, and Country	Research Population	*N*	Results		
			Factor Structure	**Internal Consistency**	**Test-Retest Reliability**
			with 14 items (less item 14) – RMSEA = .025, CFI = .961, TLI = .944	Factor 2 to Factor 3: r_s = .24	
Zuaboni et al., 2021 (Switzerland & Germany)	General psychiatric inpatient ward staff	392	EFA, CFA: 3 factors, with 15 items – SRMR =.05, RMSEA = .04, CFI = .92	Total: α = .74, Factor 1: α = .62, Factor 2: α = .55, Factor 3: α = .69	

Note. α = Cronbach's alpha, CFA = Confirmatory Factor Analysis, CFI = Comparative Fit Index, ECV = Explained Common Variance**,** EFA = Exploratory Factor Analysis, ESEM = Exploratory Structural Equation Modelling, ICC = Intraclass Correlation Coefficient, PCA = Principal Component Analysis, PSI = Person Separation Index, PUC = Percent of Uncontaminated Correlations, *r* = Pearson correlation coefficient, r_s = Spearman's rho, RMSEA = Root Mean Square Error of Approximation, SEM = Structural Equation Modelling, SRMR = Standardized Root Mean Squared Residual, TLI = Tucker-Lewis Index, WRMR = Weighted Root Mean Squared Residual, x^2 = Chi-square. The data presented in the table are the aggregate data rather than data for each group in that study.

Recurring Themes

There are no single best measures of attitudes regarding personal recovery and mental distress stigma. The limited investigations on the translated versions of the RAQ-7 and OMS-HC reveal that both have not been extensively validated, and the findings of these international validations varied. It is difficult to compare the results of different studies due to the varying methodologies used in the studies and the lack of explanation for the chosen approach (e.g., CFA & EFA vs SEM only, internal consistency over temporal reliability, no mention of factor analytic process), metric (e.g., constant use of Cronbach's alpha as a reliability estimate), and decision-making process (e.g., whether all or some of the cutoff criteria for fit indices are to be met). The use of diverse methods and underreporting of the rationale behind the methods are some of the main issues that could prevent test users from understanding and replicating findings (see the next section for a more detailed discussion of these issues). If findings are not replicated, the validity of the measure and, subsequently, its whole foundation or support as a measure of recovery or stigma would be in question. Also, it would be difficult to culturally-adapt measures to a target population. Because measurements are used to establish outcomes and guide quality improvement efforts in mental health care, the goal of promoting stigma-free and recovery-oriented individuals, professionals, and systems may, therefore, be hindered. Further validation of these measures of attitudes is needed to ensure that they are appropriate in assessing interventions that support policy mandates.

Issues and Recommendations in Measurement

Practices and Equivocal Outcomes

A key issue highlighted in the earlier review of validation literature is the use of diverse non-transparent measurement practices which may have resulted in partial support for the psychometric properties of the RAQ-7 and OMS-HC. While every researcher has the right to choose how to approach a research problem or arrive at an evidence of validity for this context, underreporting diverse measurement practices does not facilitate understanding and support communication of validation findings, given that not all research users are knowledgeable in psychometrics (Hussey & Hughes, 2020). Flexibility could be problematic when transparency is not maintained, as it cannot thwart ignorance, negligence, and misrepresentation of measurement practices (Flake & Fried, 2020). Underreporting is an example of a *questionable measurement practice*, which Flake and Fried (2020, p. 457) define as "decisions researchers make that raise doubts about the validity of measure use in a study, and ultimately the study's final conclusions". In other words, when crucial measurement information is omitted in research papers, such as the earlier mentioned inference criteria for interpreting fit indices in the RAQ-7 and OMS-HC validation studies, it is not always possible to verify the claims made by the authors. Some of the past validators have considered satisfying one or a few fit index cutoff(s) as the basis for statements that the data supported the known structures and factorial validity of these measures. Further, when these measures with equivocal validity support are used in another study, as in evaluating anti-stigma and recovery-oriented intervention, the overall validity of a conclusion from that study could not be assessed (Flake et al., 2017).

The issue of the methodological diversity and underreporting employed by various validators that may have led to equivocal evidence of validity for the RAQ-7 and OMS-HC is

similar to the proceeding dilemma in Hussey and Hughes (2020). Despite good internal consistency and temporal reliability, they found that only 73% of the social psychology measures they had studied had good factorial validity (i.e., fit indices meet the cutoff values - SRMR ≤ .09, RMSEA ≤ .06, CFI ≥ .95, and TLI ≥ .95). When the same principles in assessing factorial validity are applied to the validation findings for the two measures under study, only one of the six confirmatory studies has data with a good fit to the known three-factor structure for the OMS-HC-15. In contrast, neither of the two confirmatory studies has data that provide unequivocal support to the known two-factor structure for the RAQ-7. With the similarity of these findings, it can be questioned whether the same conclusion in Hussey and Hughes (2020) could be made to the validation literature of the RAQ-7 and OMS-HC. Could the underreporting of certain important psychometric information, mean that evidence of the invalidity of the two measures is hiding in plain sight? There seems to be the possibility that specific goodness of fit indices was used to compare models and increase apparent factorial validity, consistent with findings on how there is inadequate proof that measures accurately assess the constructs intended to be measured (Barry et al., 2014). Also, validators may have cherry-picked facts to support their psychometric constructs to achieve specific academic goals, similar to how unreasonable measurement flexibility prevails as researchers conveniently select parts of questionnaires that best suit their purpose (Orben & Przybylski, 2019).

As a partial solution to the highlighted issue, Hughes and Hussey (2020) advised that researchers should collaborate in an inclusive process to develop and evaluate the psychometric properties of a measure. They also suggested that the preregistration of methodologies could help limit the flexibility of researchers, such as the conduct of unplanned spurious analyses. Preregistration can help prevent or minimise cherry-picking of validity information and analyses,

as hypotheses and analytical procedures are already defined before data analysis (Yamada, 2018). It was further recommended that sharing data and materials (e.g. statistical program syntax or code) with the public within ethical guidelines will permit replication, extension, and meta-analytic validation and encourage checks and balances (Asendorpf et al., 2016; Munafò et al., 2017). Thus, open and robust principles such as using preregistered inference criteria for interpreting model-data fit index cutoffs are recommended so that consumers of findings have a consistent guideline for interpreting support for known or competing models. Such additional yet crucial measurement information could make further subsequent replication or extension of past works on the validity of a measure easier to perform.

Practices and Lesser Research Interest

The poor or mixed validity evidence in the validation literature for both the RAQ-7 and OMS-HC could not be the only unfavourable outcome of diverse non-transparent measurement practices. The dearth of validation studies could also be brought about by these practices, congruous to how lack of transparency in data gathering and processing hinders the ability to evaluate and reproduce others' work as there is no visibility of what would qualify as "good" or "problematic" (Chambers, 2017; Simmons et al., 2011). Potential researchers wanting to validate the two measures may have been disheartened to pursue their study as they found it challenging to understand the earlier validation process and findings. For example, anyone wishing to replicate or extend the validity evidence of the RAQ-7 may find it confusing which factor analytic approach would be followed between the confirmatory factor analyses only of Chiba et al. (2016) and the exploratory and confirmatory factor analyses of Wilrycx et al. (2012). In the case of the OMS-HC, anyone wanting to refine the number of items needed for a more stable three-factor structure may struggle with which combination of absolute and incremental fit

indices would be used, considering that past validators claimed that their data, analysed through their set of fit indices, fitted to the known structure. This measurement practice issue resulting in equivocal outcomes and subsequent lesser research interest parallels with Flake et al. (2017). They revealed that underreporting robust metrics, misuse and abuse of Cronbach's α, and insufficient validation in the different construct validation phases are prevalent within research works for well-known measures. They concluded that such diverse non-transparent measurement practices make problematic conclusions less appetising for subsequent research to replicate findings because there is no sufficient and precise information to work with beforehand.

Fortunately, several best practices can provide partial, if not complete, solutions to the issue of diverse non-transparent measurement practices that decrease the chance of subsequent research to replicate findings and stimulate further research interest. It has been suggested that validators or researchers should shift from focusing on performing the traditional modal practice of assessing the internal consistency of scores only to also estimating whether scores are stable over time (i.e., temporal reliability) and investigating whether the scores support the structure of the latent constructs within the measure (i.e., factorial validity, Putnick & Bornstein, 2016). In other words, construct validation, which is the consolidation of several pieces of information to establish the reason behind a number chosen to exemplify a psychological concept (Cronbach & Meehl, 1955), must be well-thought-out and performed following the three phases of validation by Flake et al. (2017). Phase one, the *substantial* phase, entails defining the theoretical origins, characteristics and nomological network of the construct through a thorough literature review. Phase two, the *structural* phase, examines the form of the construct quantitatively, where empirical analyses are undertaken to see if the results match the expected pattern, whether the content is verifiable, and whether the measure holds up over time across the groups. The final is

the *external* phase, where convergent, discriminant, and known-group validities are examined. Therefore, when the above recommendations are applied, dubious conclusions obtained in the external phase would not have been drawn and generated a lesser appetite for further research because problems at the substantive (e.g., the known three-factor model for the OMS-HC had already had a solid theoretical and empirical support) and the structural phase (e.g. RAQ-7 scores have been demonstrated to be internally and temporally consistent) were already identified and resolved. Suppose the different phases of content validation are performed with care, anyone proposing to assess, say, an anti-stigma and recovery-oriented intervention using the measures under study or further investigate the psychometric properties of these measures will have clear and sufficient information to do so.

Practices and Limited Evidence of Validity

Another potential outcome of diverse non-open measurement practices is the state of having limited validity evidence for the two measures under study. Only two studies have investigated the factorial validity of the RAQ-7, despite it existing for more than two decades now. Only a few validators have endeavoured to test alternative structures for the OMS-HC (i.e., Happell et al., 2019; Őri et al., 2020), despite some studies that concluded the known three-factor structure had a good fit based on specific met inference criteria only. While doing more confirmation for the known structure is not necessarily problematic, given that there are only a few done for the RAQ-7 and the original English version of the OMS-HC, the exclusion of testing other competing measurement models does not provide adequate support for the construct validity of a measure (DiStefano & Hess, 2005). Such practice does not offer sufficient information on the performance of measures and does not help operationalised measures to be further validated, consistent with the conclusion made by Crutzen and Peters (2017) in their

review of measures in health psychology. They found that several researchers describe the psychometric properties of the measures they use through Cronbach's α only. Further, only 3% of the studies they investigated demonstrated evidence of factorial validity, despite that it is a vital feature to ensure a measure is valuable (Gerbing & Anderson, 1988). Therefore, the lack of factor analyses within the validation literature for the RAQ-7 and OMS-HC means that these measures are not structurally valid, given that the least common validity evidence of a measure was reported in the literature, the more likely the measure would not demonstrate that validity evidence (Hussey & Hughes, 2020). However, summing scales is more commonplace in psychology than analysing latent variables, as researchers often add (or average) numerical values of responses from measures (Bauer & Curran, 2016), which is another measurement practice issue that does not offer further support to the validity of a measure.

Given the multidimensionality of the RAQ-7 and OMS-HC, it may be argued that a unidimensionality test is no longer required. However, recent recommendations calling for more robust validation of measures have posited that if the unidimensionality of measures is not assessed, there is a possibility that a measure will have items that reflect not only the different dimensions of the construct under study but also that of other constructs (Ziegler & Hagemann, 2015). If items in a measure are not determined as unidimensional, test scores cannot be appropriately interpreted, including whether subsets of several scores should not be calculated (Revelle & Zinbarg, 2009; Stout, 1987). Further, those who use the measure cannot qualify their interpretations regarding whether they require information about one, two or more dimensions of the construct (McNeish & Wolf, 2020). Therefore, the practice of testing unidimensional models is important as it permits the gathering of solid evidence of an effect across settings, aligned with the notion that further studies or replication of a previous finding or proof of validity help in both

reducing the propagation of erroneous statements and maximising resources (Nosek et al., 2012; Sijtsma, 2016). When there are more confirmations that the RAQ-7 and OMS-HC are either unidimensional or not, users of these measures will be able to have more information that could guide them on how they could interpret and use the scores and more time to be allotted to exploring other under-optimised psychometric properties and alternative measurement models.

While there were findings that depicted the RAQ-7 as unidimensional (i.e., Hungerford et al., 2015; Wilrycx et al., 2012), it has been recognised that personal recovery, the core construct measured by the RAQ-7, is multidimensional (van Weeghel et al., 2019). Similarly, it has been noted that the OMS-HC is not single-factored (e.g., van der Maas et al., 2018) and that mental distress stigma is a complex construct (Fox et al., 2018), whereas there were findings not supporting the known three-factor model (e.g., Őri et al., 2020). The preceding equivocations raise the question of whether the known factor structure and unidimensional model of a measure are sufficient to offer evidence for the construct validity of that measure. Such equivocations are similar to the predicament that, while self-report measures are predominantly designed to assess just one construct (Reise & Haviland, 2005), constructs assessed by any given measure are inherently multifaceted (Chen et al., 2006). Fortunately, there already have been efforts to help validators with the difficult task of examining one construct while also observing multiple facets of it. As a partial solution to this problem, it has been recommended that bifactor models be used when most items of a measure have been classified into one factor in a previously used model but with a few falling into a second or third factor to eliminate partially overlapping dimensions (Reise et al., 2007). In other words, a bifactor model could help to address how unidimensional or multidimensional a measure is (Hammer & Toland, 2016), as each observable variable (test question or item) is allowed to contribute to two latent variables, allowing for a more significant

correlation of items to multiple sources of variance (Morin et al., 2016). The first latent variable reflects a general factor at the core of all the scale items, and the second is a specific factor (Boateng et al., 2018; Gerbing & Hamilton, 1996). In the OMS-HC validation literature, Őri et al. (2020) leveraged the bifactor model to offer an alternative explanation of the internal structure of the measure. Their bifactor analysis emphasised that such a model can allow the test user to measure the overall stigmatising attitudes and behavioural intent when a single score is needed to assess an anti-stigma intervention. Also, if a researcher needs to quantify a factor of mental distress stigma, say, social distance, the bifactor model could permit the researcher to use the subscale score representing that factor.

Building upon the recommendations in the previous sections, future studies should consider the various components of their methodologies to confirm or explore the validity of measures, such as the RAQ-7 and OMS-HC. These studies should also provide information about what and how the decisions relate to the validation of the measures. Understanding best measurement practices is necessary for making pieces of evidence possess more weight in supporting conclusions. Information about the quality of evidence enables more confidence in using these measures, potentially gaining unequivocal validity outcomes and generating further validation research interest.

The Current Study: Replicating Replications

The current study aimed to replicate prior validation studies of two measures of attitudes toward personal recovery and mental distress stigma, which are the Recovery Attitudes Questionnaire (RAQ; Borkin et al., 2000) and the Opening Minds Scale for Health Care Providers (OMS-HC; Kassam et al., 2012; Modgill et al., 2014), respectively. It was important to validate these measures so that interventions and other efforts using these measures and

associated with policy mandates can be assessed accurately. The current study validated the seven-item RAQ (RAQ-7) and the fifteen-item OMS-HC (OMS-HC -15, including the OMS-HC-14), which were the versions that the authors of these measures eventually generated and suggested to be used after their development and initial evaluation processes. Although the evidence of the validity of these two measures remains equivocal due to the measurement practices used in previous studies, these versions were the most frequently validated. The current study specifically tested the significant findings on the factorial validity, internal consistency and test-retest reliability of the RAQ by Chiba et al. (2016) and the OMS-HC by Őri et al. (2020). Further, the current study partially replicated the replications in validating the RAQ and OMS-HC instead of basing the validation on the broader literature review. Lastly, as the current study is not the first to reveal the measure structures, it was prudent to confirm rather than explore further.

Recovery Attitudes Questionnaire

Chiba et al. (2016) was the only group to have undertaken a comprehensive structural validation of the RAQ-7 following the work conducted by the authors of the RAQ. They examined the factorial validity (through confirmatory analyses), internal consistency, test-retest reliability, and measurement invariance of the measure. Further, Chiba et al. (2016) tested the unidimensionality of the RAQ-7, which is important given the multidimensional nature of recovery. While Wilrycx et al. (2012) have also explored this unidimensionality (through EFA), the approach of Chiba et al. (2016) was more aligned with issues about overfitting risks (Fokkema & Greiff, 2017) and not doing exploratory and confirmatory analyses on the same dataset (Ziegler, 2014). However, there are methodological limitations in Chiba et al.'s (2016) study. There was no appreciation of a priori reasons to use: 1) metrics such as McDonald's

(1999) coefficient omega; and 2) decision-making processes like Hussey and Hughes' (2020) approach to determining a "good", "poor", and "mixed" fit, among others. Firstly, using McDonald's (1999) coefficient omega as a reliability estimate is essential, considering that recovery and the RAQ-7 are multidimensional, the RAQ-7 is known to suffer poor internal consistency, and its factor loadings are known to be unequal. McDonald's (1999) coefficient omega must be used over Cronbach's α, which assumes that all items represent latent variables equally and that a single factor constitutes a structure of a measure (Schmitt, 1996). Secondly, as the equivocal model fit was precise in Chiba et al.'s (2016) study, a research user of their finding could better understand the factorial validity if there was a clear guideline in interpreting a model fit, like that in Hussey and Hughes (2020). For these reasons and the lack of publications about the psychometric properties of the RAQ-7, the current study intended to replicate Chiba et al.'s (2016) results, but with the guidance of open and rigorous psychometric research practices.

Opening Minds Scale for Health Care Providers

Given that bifactor modelling complements unidimensional and correlated factor analyses (Reise et al., 2010), the present study replicated Őri et al.'s (2020) research to better understand the internal structure of the OMS-HC. This replication provided evidence for the unidimensional, correlated factors and bifactor models. It helped clarify which model best describes the factor structure of the measure and which version (in terms of items) could be most suitable. This replication aimed to provide evidence on the "shared conceptual theme" notion that Modgill et al. (2014, p. 8) and Őri et al. (2020, p. 7) have investigated, where there were strong and significant correlations between the specific factors and the general factor and weak yet significant correlations between the specific factors. However, like Chiba et al. (2016), several methodologies were limited in Őri et al.'s (2020) study. Őri et al. (2020) did not employ a stable

absolute fit index like the standardised root mean square residual in their confirmatory analyses (Hu & Bentler, 1999; Hussey & Hughes, 2020). Therefore, the current study endeavoured to replicate Őri et al. (2020) through a preregistered approach and with the guidance of best practices in inspecting a range of psychometric properties.

Summary

The current study's replication of Chiba et al. (2016) and Őri et al. (2020) was non-exact. The exploratory analyses in Őri et al. (2020) and the interitem reliability estimation in Chiba et al. (2016) were excluded. Many of the metrics, cutoffs and decision-making processes discussed in Hussey and Hughes (2020) were implemented in the current study to draw together the validation of the two measures and enable the evaluation of shortened versions. The current study was preregistered, including hypotheses, methodologies, metrics, cutoffs, decision-making processes and R script or code for data analysis. The preregistration (see https://osf.io/tuh49/) was written in the future tense as it was placed in the Open Science Framework (Foster & Deardorff, 2017) repository before data collection. The preregistration content is included in the Methods chapter.

Hypotheses

The study hypotheses were developed from the conclusions of Chiba et al. (2016) and Őri et al. (2020) but were not the verbatim implications of the prior validation studies.

Recovery Attitudes Questionnaire

The present study fitted first-order CFA models for the RAQ-7: one was the two correlated factors –1) *"Recovery is possible and needs faith,"* and 2) *"Recovery is difficult and differs among people"* - that covary with each other, which was the best fitting model in Chiba et

al. (2016), and the second was the unidimensional model. The two-factor structure fitted in Chiba et al. (2016) satisfied all the absolute fit indices cutoffs but not the comparative fit index (CFI = .86). Although the inference criteria used by Chiba et al. (2016) were not consistent with the widely used approach of Hu and Bentler (1999), their negative result on the supplemental fit index CFI reflected that the overall fit of the model was somewhat equivocal. It was hypothesised, therefore, that this model would exhibit a "mixed fit" (*Hypothesis 1a*) to the current data but would be better than the unidimensional model (*Hypothesis 1b*).

Further, it can be noted that the factor loading of item 2, "To recover requires faith", on the two-factor structure was small (.26) in Chiba et al. (2016), consistent with Hungerford et al. (2015) and Wilrycx et al. (2012). It was hypothesised that the factor loading of item 2 would not exceed .3 (*Hypothesis 1c*).

While Chiba et al. (2016) considered that evidence of factorial validity was acceptable, they noted that the RAQ-7 suffered from having a small number of items. They found that the RAQ-7 had unsatisfactory internal consistency coefficients (α = .53 to .64 for the total and factor scores) and moderate total score test-retest reliability (intraclass correlation coefficient [ICC] of .68). Therefore, the following was hypothesised in the present study: the 95% confidence interval lower bound of the total and subscale RAQ-7 scores' internal consistency will be $\omega_t < .7$ (*Hypothesis 1d*); and b) the 95% confidence interval lower bound of the RAQ-7 total and subscale score test-retest reliability will be $r < .7$ (*Hypothesis 1e*).

Opening Minds Scale for Health Care Providers

The present study fitted three CFA models; that is, a single-factor model, the three correlated factors of Modgill et al. (2014; 15 items) – 1) *Attitudes,* 2) *Disclosure and Help-*

seeking, and 3) *Social Distance*, and the bifactor solution of Őri et al. (2020), with the observed variables loading to both a general factor called "overall stigmatising attitude" (p. 5) and to the three factors of Modgill et al. (2014). Őri et al. (2020) found that their structure is the best fitting model based on supplemental fit indices, while the three-factor structure was not in the acceptable range. All models examined in Őri et al. (2020) had normed chi-square values below 2, which were not within their inference criteria (i.e., 2-5). Nevertheless, a ratio of ≤2 indicates a superior fit (Cole, 1987). Therefore, it was hypothesised in the present study that the bifactor solution will exhibit a "good" fit to the current data (*Hypothesis 2a*) and will be superior to the three-factor structure (*Hypothesis 2b*). Consequently, it was hypothesised that the three-factor structure (*Hypothesis 2c*) and the single-factor structure (*Hypothesis 2d*) would exhibit a "poor" fit.

Őri et al. (2020) claimed that their bifactor solution is another option in explaining the internal structure of the OMS-HC but found it to be deficient in model-based reliability, with .71 percent of uncontaminated correlations (PUC). The general factor of the bifactor model displayed minimum acceptable internal consistency reliability (coefficient omega hierarchical of .56), but the reliability of the three specific factors was insufficient. The present study aimed to determine if these hold with another data set. The following were hypothesised: the bifactor model's PUC < .8, ECV > .6, and ω_h >.7 (*Hypothesis 2e*); the 95% confidence interval lower bound of the total OMS-HC-14 (*Hypothesis 2f*) and OMS-HC-15 (*Hypothesis 2g*) scores internal consistency would be $\omega_t \geq .7$; and, the 95% confidence interval lower bound of the OMS-HC-14 (*Hypothesis 2h*) and OMS-HC-15 (*Hypothesis 2i*) subscales scores' internal consistency would be $\omega_t < .7$.

Further, the test-retest reliability coefficients of the general (ICC = .95) and specific factors (ICC from .84 to .90) were good to excellent (Őri et al., 2020). Therefore, it was hypothesised that the 95% confidence interval lower bound of the OMS-HC-14 (*Hypothesis 2j*) and the OMS-HC-15 (*Hypothesis 2k*) the total and the subscale scores test-retest reliability would be $r \geq .7$.

Lastly, Őri et al. (2020, p. 7) supported the notion of a "shared conceptual theme" of Modgill et al. (2014, p. 8), where there were strong and significant correlations between the specific factors and the general factor and weak yet significant correlations between the specific factors. Therefore, it was hypothesised that: the correlations between the general and specific factors would be $r_s > .59$ at $\rho > .05$ *(Hypothesis 2l)*. However, the correlations between the specific factors would be $r_s \leq .39$ at $\rho > .05$ (*Hypothesis 2m*).

The hypotheses of the current study were enumerated again below to facilitate reference.

Recovery Attitudes Questionnaire.

- *Hypothesis 1a:* the two-factor model for the RAQ-7 would exhibit a "mixed fit" to the current data;
- *Hypothesis 1b*: the two-factor model for the RAQ-7 would exhibit abetter fit than the unidimensional model;
- *Hypothesis 1c*: the factor loading of item 2 on the two-factor structure would not exceed .3;
- *Hypothesis 1d*: the 95% confidence interval lower bound of the total and subscale RAQ-7 scores' internal consistency would be $\omega_t < .7$;

- *Hypothesis 1e*: the 95% confidence interval lower bound of the RAQ-7 total and subscale score test-retest reliability would be $r < .7$.

Opening Minds Scale for Health Care Providers.

- *Hypothesis 2a*: the bifactor solution would exhibit a "good" fit to the current data;
- *Hypothesis 2b*: the bifactor solution would exhibit a superior fit to the current data than the three-factor structure;
- *Hypothesis 2c*: the three-factor structure would exhibit a "poor" fit;
- *Hypothesis 2d*: the single-factor structure would exhibit a "poor" fit;
- *Hypothesis 2e*: the bifactor model's PUC < .8, ECV > .6, and ω_h >.7;
- *Hypothesis 2f*: the 95% confidence interval lower bound of the total OMS-HC-14 scores internal consistency will be $\omega_t \geq .7$;
- *Hypothesis 2g*: the 95% confidence interval lower bound of the total OMS-HC-15 scores internal consistency will be $\omega_t \geq .7$;
- *Hypothesis 2h:* the 95% confidence interval lower bound of the OMS-HC-14 subscales scores internal consistency will be $\omega_t < .7$:
- *Hypothesis 2i:* the 95% confidence interval lower bound of the OMS-HC-15 subscales scores internal consistency will be $\omega_t < .7$;
- *Hypothesis 2j*: the 95% confidence interval lower bound of the OMS-HC-14 total and subscale scores test-retest reliability will be $r \geq .7$;
- *Hypothesis 2k*: the 95% confidence interval lower bound of the OMS-HC-15 total and subscale scores test-retest reliability will be $r \geq .7$;
- *Hypothesis 2l:* the correlations between the general and specific factors will be $r_s >$ .59 at $\rho > .05$;

- *Hypothesis 2m:* the correlations between the specific factors will be $r_s \leq .39$ at $\rho > .05$.

Chapter 2: Methods

Design

This study was a cross-sectional (observational) study. No manipulation by way of blinding was done.

Participants

Participants were recruited from Prolific Academic to complete an online Qualtrics survey. Prolific Academic is a crowdsourcing platform that has survey respondents across the world (albeit predominantly in the United States and the United Kingdom; Prolific, 2022a). The Prolific Academic participants in the study had an approval rating of 95% or better, were fluent in the English language and had their employment sector tagged under 'Medicine'. The said approval rating was to ensure participants that were recruited are attentive, honest and reliable, consistent with the research on the association between the use of such prescreeners and high overall data quality (Eyal et al., 2021). The English fluency prescreener was used to ensure that participants will have understood and followed the instructions of the study and the two measures being validated, which were both in the English language, consistent with Prolific Academic's (2022b) recommendations. Lastly, the 'Medicine' prescreener, which is the same data filter to capture Prolific Academic participants working in health care (Prolific, 2022c) was used given the aim of the current study to validate the RAQ-7 and OMS-HC with health care providers. A total of 286 participants completed the survey after 34 participants were removed based on data exclusion criteria (see Data Analysis). Nineteen of the participants completed the two measures again after two full calendar weeks.

Measures

Participants completed an online Qualtrics survey (see Appendix A). The survey started with three items to gather demographic information to describe the participants. Direct questions were used to gather the following information: age (17 or younger, 18-20, 21-30, 31-40, 41-50, 51-60, 61 or older), gender (male, female, non-binary, prefer not to say, and other), and ethnicity (see page 149 in Appendix A for the list of response options). These demographic data were reported but were not used in the primary data analyses. Following the demographic information items, participants completed the RAQ-7 (Borkin et al., 2000; see page 150 in Appendix A) and the OMS-HC-15 (Kassam et al., 2012; Modgill et al., 2014; see pages 151-152 in Appendix A), which were described in Chapter 1. The English stems in the scales were presented verbatim as in the development and evaluation studies.

Recovery Attitudes Questionnaire

The RAQ-7 is a seven-item self-report measure of attitude toward mental health recovery that was developed with consumers (Borkin et al., 2000). The items are known to load to two correlated factors – 1) *"Recovery is possible and needs faith,"* and 2) *"Recovery is difficult and differs among people".* A 5-point Likert scale response options was used for each item, ranging from Strongly Disagree (1) to Strongly Agree (5). Total and subscale scores were computed by summing the item ratings. The possible score range is 7-35 for the total scale, 4-20 for the factor 1 subscale (items 2, 4, 5 and 6) and 3-15 for the factor 2 subscale (items 1, 3, and 7). Higher scores indicate more positive attitudes about recovery.

Opening Minds Scale for Health Providers

The OMS-HC-15 is a fifteen-item self-report measure of attitudes and behavioural intentions towards people with lived experience of mental distress (Kassam et al., 2012; Modgill et al., 2014). The OMS-HC-15 items are known to load to three correlated factors – 1) *Attitude*, 2) *Disclosure and Help-seeking*, and 3) *Social Distance.* The items are rated on a 5-point Likert scale from Strongly Disagree to (1) to Strongly Agree (5). Items 3, 8, 9, 10, and 19 were reverse scored. Total and subscale scores were created by adding each response for the items within each scale. The possible score range is 15-75 for the total scale, 6-30 for the factor 1 scale (items 1, 12, 13, 14, 18 and 20), 4-20 for the factor 2 scale (items 4, 6 and 10) and 5-25 for the factor 3 scale (items 3, 8, 9, 17 and 19). Lower scores denote less stigmatising attitudes and behavioural intentions.

Procedure

Participants were recruited from Prolific Academic, the crowdsourcing platform where the current study was also advertised. Participants were provided with key information to help them decide if they wished to participate. Those who expressed interest to participate were invited to click the link that took them to the online Qualtrics survey. On the study link, participants were further informed about the details of the study (see Information Sheet in Appendix A, pages 145-146). Participants were eligible to participate if they responded "Yes" to the study consent and the worker in the 'Medicine' sector questions and were at least 18 years old. Otherwise, respondents were exited from the survey, and their replies were disregarded. At the end of the survey, participants were informed how they confirm that they have completed the study. Participants were paid GBP 1.25 for their participation in the survey.

Data collection ceased after 286 participants answered, which came before the other set stopping rule of two full calendar weeks. This sample size was 10% more than planned (see Sample Size) and was specified in anticipation that some participants may have finished the questionnaires but their data may not be used because of the exclusion criteria (see Data Analysis). After two calendar weeks, a subsample of the participants (n = 19; 10% higher than the 17, to account for possible data exclusion) was randomly selected through the fishbowl method (Sevilla et al., 1993) technique and re-surveyed for test-retest reliability estimation. This 2-week interval was similar to the follow-up period of Chiba et al. (2016). Data collection ceased after 19 participants responded, which came first before the other set stopping rule of two full calendar weeks. The test-retest reliability analysis would have been abandoned if the target sample size was not achieved within the planned timeframe. The "Prevent multiple submissions" option in Qualtrics was turned on to prevent participants from taking the survey more than once but was turned off before the retest.

Ethics

The Massey University Human Ethics Committee was notified (Ethics Notification Number: 4000025270; see Appendix B) that the current study was assessed as low risk. Participants were informed about the details of the study and offered the option of joining or not. An amount that could not be used as an inducement was paid to the participants in exchange for the time spent on completing the measures. Additionally, participants were informed before the survey that their data would be anonymised and stored in an open internet repository (i.e., the Open Science Framework; Foster & Deardorff, 2017). While it was not expected that the two measures would cause psychological harm, a link to information regarding psychological support and services was provided at the end of the survey.

Data Analysis

Participants who answered fewer than 80% of the 22 questions in the two measures (i.e., missed more than four items), similar to the criteria set for validating the OMS-HC-15 (i.e., van der Maas et al., 2018), were excluded during data processing. Towards the end of the survey, respondents were prompted to tick "Agree" to an attention check question. Participants who answered this question with a response other than "Agree" or did not respond were omitted from the data processing. Upon applying these exclusion criteria, if there were duplicate responses from the same Prolific participant, only the most recent response was retained, and the others were excluded.

Although not anticipated to be prominently used, single imputation was employed. The R software (version 4.1.0; R Core Team, 2021), the amelia package (Honaker et al., 2011) and the mice package (van Buuren & Groothuis-Oudshoorn, 2011) were used for such imputation. Imputation was required for the CFA (even if it used a different estimation method).

Confirmatory factor analyses (CFA) were employed to evaluate the model fit indices of the known and competing structures for the RAQ-7 and the OMS-HC. McDonald's (1999) omega total and Pearson's correlation coefficient were used for internal consistency and test-retest reliability estimation, respectively. Lastly, Spearman's correlation coefficient was used to examine the notion of a "shared conceptual theme" (Modgill et al., 2014, p. 8).

Sample Size

Sample Size to Detect Model Misspecification. Several methods to determine the sample size associated with sufficient power to detect the misspecification of a model were considered. A sample size of 260 was more than enough to reject an incorrect model for the two-

factor structure of the RAQ-7, with 17 parameters (see RAQ-7 Models in Model Identification and Specification) and a 15:1 sample size to estimated parameter ratio[2], and the bifactor solution of the OMS-HC with 52 parameters (see OMS-HC Models in Model Identification and Specification) and a 5:1 ratio. These ratios were greater than or equal to the recommendation of 5:1 by Bentler and Chou (1987) and aligned with more recent simulation studies that recommend 30 samples for a simple CFA with four indicators (Wolf et al., 2013) and 50 to 70 samples involving four latent variables (Sideridis et al., 2014).

In terms of empirically supported power analysis using CFI (Kim, 2005), the sample size of 260 was more than enough to yield 90% power for examining the two-factor structure for the RAQ-7 at $\rho \leq .05$. In computing the former, the expected CFI value (i.e., .91) used was just above Hu and Bentler's (1999) cutoff. The number of items used in the calculation was 8, as the online calculator works with the number of items as multiple of the number of the factors (Arifin, 2018). If the number of items entered was 6, the sample size of 260 is still more than sufficient to yield the same study power obtained with 8 items. Further, the average factor loading was .7, which is in line with known cutoff benchmarks (e.g., MacCallum et al., 2001; Tabachnick et al., 2007) and just about the average in Borkin et al. (2000), where the English version of the instrument was used. The average factor correlation was set to default (i.e., .3; Arifin, 2018) and the expected dropout rate to zero.

[2] The sample size to estimated parameter ratio is the number of observations or participants for each estimated parameter in the model (Jackson, 2003). As an example, for the two-factor structure for the RAQ-7, the 15:1 ratio is achieved by dividing the sample size (i.e., 270) by the number of parameters (i.e., 17).

Using Kim's (2005) method with root mean square error of approximation (RMSEA), a sample size of 260 was more than enough to produce 90% power to detect misspecification of the OMS-HC bifactor solution at $\rho \leq .05$. The RMSEA value (i.e., .045) used for computing the said power was similar to the replication of Őri et al. (2020) for the three-factor structure of Modgill et al. (2014). The number of factors entered was 4 to represent the bifactor solution, the best-fitting model in Öri et al. (2020), where items are expected to load to the "overall stigmatising attitude" (general factor; p. 5) alongside three correlated factors (or specific factors) of Modgill et al. (2014). The number of items was 14 to reflect the OMS-HC version used. The expected dropout rate was set to zero.

The sample size of 260 was less than the 307 in Chiba et al. (2016) but greater than the 211 in Őri et al. (2020). However, there was no compelling reason to justify the financial investment associated with recruiting a sample size greater than that in Chiba et al. (2016), given that 260 respondents would be adequate.

Sample Size to Detect Target Parameter. For the internal consistency of the RAQ-7, the sample size of 260 was more than enough to have a confidence interval not wider than .05 more than 85% of the time for the population value of coefficient omega. This computation was done using the accuracy in parameter estimation (AIPE) for the reliability coefficients approach (Maxwell et al., 2008), the R software (version 4.1.0; R Core Team, 2021) and the MBESS R package (version 4.8.1; Kelley, 2017). The factor loadings and error variance values in Chiba et al. (2016) were employed to plan this sample size. As the required information was not readily available in Őri et al. (2020), this sample size calculation was not done for the OMS-HC.

Machin et al.'s (2009) method was utilised to compute the test-retest reliability estimation sample size. For the test-retest reliabilities of both measures, a sample size of 17 was required for

90% power to detect a Pearson's correlation value of .7. Arifin's (2018) web calculator was used to compute the sample size needed to examine the two measures' factor structures and test-retest reliability.

Confirmatory Factor Analyses

CFAs were completed to examine the model that best explains the factor structures of the RAQ-7 and the OMS-HC. CFAs were conducted using the R software (version 4.1.0; R Core Team, 2021) and the lavaan package (version 0.6-9; Rosseel, 2012). For all the CFAs, maximum likelihood estimation with robust standard errors and a mean- and variance-adjusted approach (MLMV) was used, similar to Ori et al. (2020) but not Chiba et al. (2016). Rhemtulla et al. (2012) recommended that a robust maximum likelihood estimation be employed in structural equation modelling measures with five response options, as in the case of the RAQ and OMS-HC. Therefore, in line with MLMV, the present study assumed the measures' items as continuous and that the multivariate data was non-normal. Mardia's (1970) multivariate skew and kurtosis tests using the psych package (version 2.1.9; Revelle, 2021) were done, but the results were not set to change the planned CFA. This decision was based on the recommendation of Hussey and Hughes (2020) and Flake and Fried (2020) that preregistering methodologies and complying with them could help limit the flexibility that leads to the conduct of unplanned spurious analyses, which contributes to equivocal findings. Lastly, the complete standardised solution was used in the CFAs, where factor loadings were standardised by the standard deviation of both the factor and the item.

Model Identification and Specification. Five models were fitted to the current data - two models for the RAQ-7 and three for the OMS-HC.

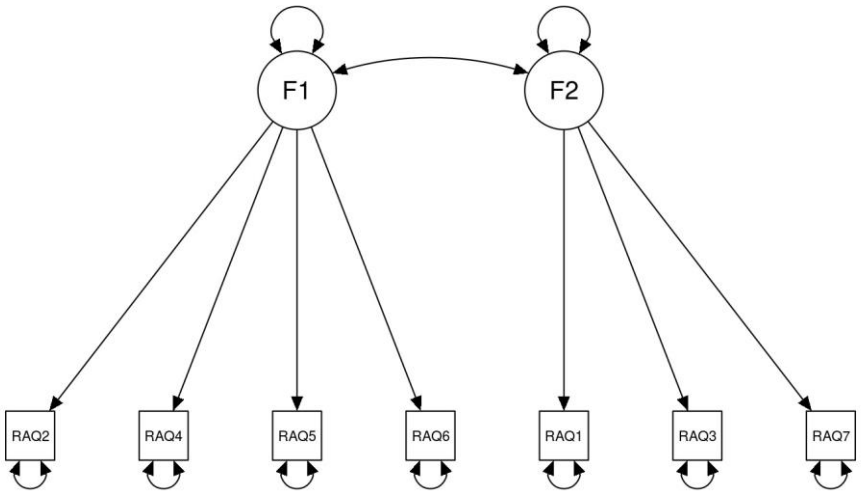
RAQ-7 Models. The two-factor structure of Borkin et al. (2000) for the RAQ-7, which was the best-fitting model according to Chiba et al. (2016), had two correlated factors (latent variables) that covaried with each other. The model contained seven items (observed variables), 28 known values[3], 17 parameters and 11 degrees of freedom[4]. All the parameters were free - seven factor loadings, seven residual variances (one per item), two total factor variances, and one factor covariance (see Figure 1). The unidimensional model for the RAQ-7 contained the same number of items and known values but had one factor, 15 parameters and 13 degrees of freedom. It differed from the two-factor structure as it had no factor covariance (see Figure 2). As the degrees of freedom of both models were positive (i.e., over-identified), model fit can be assessed.

[3] The number of known values is calculated by using the formula of $p(p+1)/2$, where p corresponds to the measure's number of items (Brown, 2015).

[4] The degrees of freedom is computed through deducting the number of free parameters from the number of known values (Brown, 2015).

Figure 1

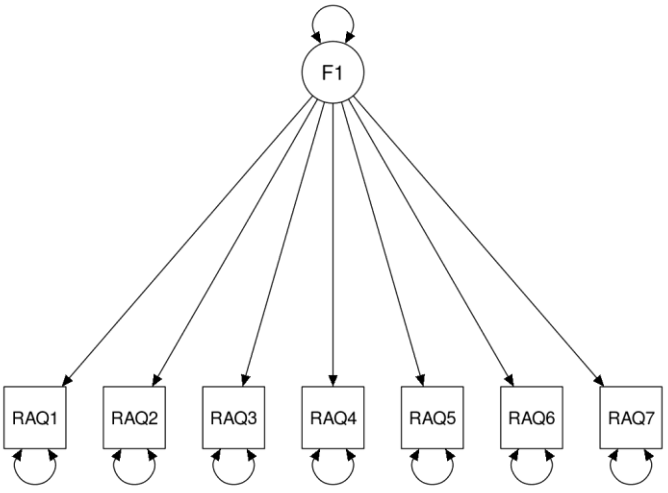
Borkin et al.'s (2000) two-factor structure for the RAQ-7.



Note. F1 = Recovery is possible and needs faith; F2 = Recovery is difficult and differs among people.

Figure 2

Unidimensional model for the RAQ-7.

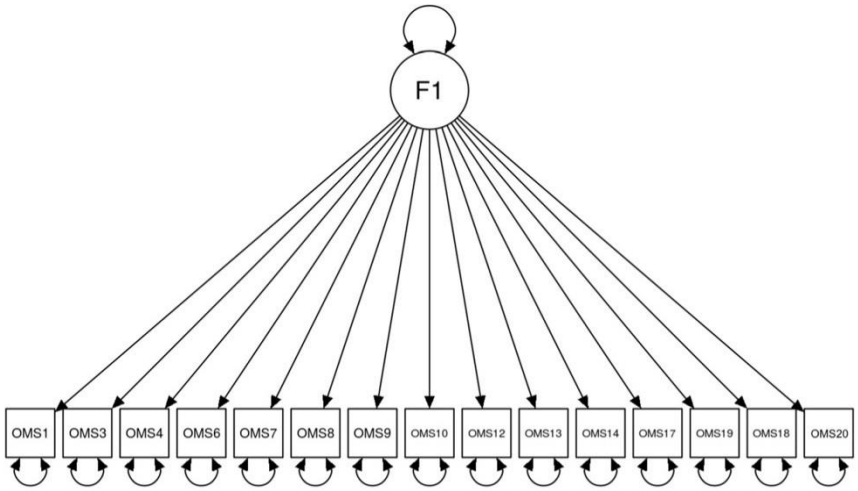


Note. F1 represents the only factor within this model.

***OMS-HC Models*.** The first model for the OMS-HC was the single-factor structure (see Figure 3), which had 15 items, 120 known values, 31 free parameters, and 89 degrees of freedom. This unidimensional model had 15 factor loadings, 15 residual variances, and one factor variance. The second was the three-factor structure of Modgill et al. (2014), where three correlated factors were allowed to covary. This model comprised 15 items, 120 known values, 36 free parameters, and 84 degrees of freedom. There were 15 factor loadings, 15 residual variances (one error term per item), three total factor variances, and three factor covariances (see Figure 4). Lastly, the best-fitting model in Öri et al. (2020) was fitted to the current data whereby there was another factor known as the "overall stigmatising attitude" (general factor; p. 5) alongside the three-factor structure of Modgill et al. (2014). The three specific factors were not allowed to covary both with other specific factors and the general factor. This bifactor solution (14 items; item 14 excluded) had each variable loaded to general and specific factors. This model had 105 known values, 52 parameters, and 53 degrees of freedom. There were 28 factor loadings, 14 residual variances (one per item), four total factor variances, and six factor covariances (see Figure 5). The six factor covariances restricted to zero to reflect the factors' orthogonality were still considered free parameters. Again, given the positive values of the degrees of freedom of each model, model fit can be assessed.

Figure 3

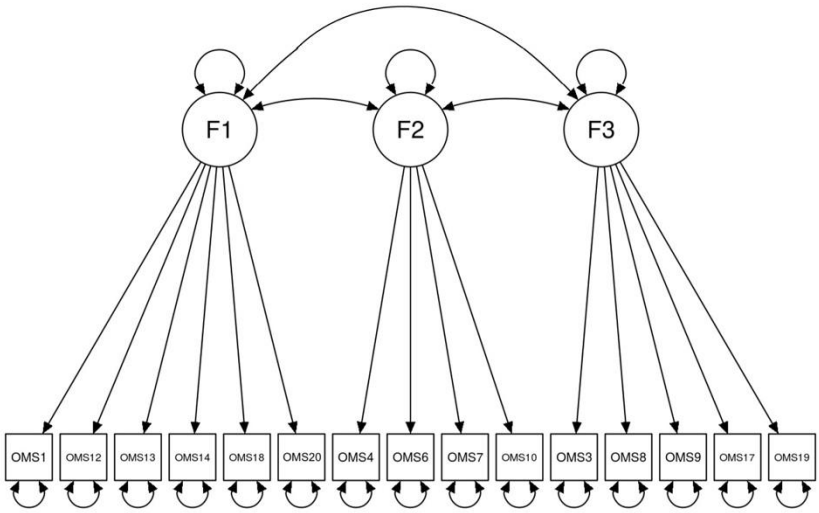
Unidimensional model for the OMS-HC-15.



Note. F1 represents the only factor within this model.

Figure 4

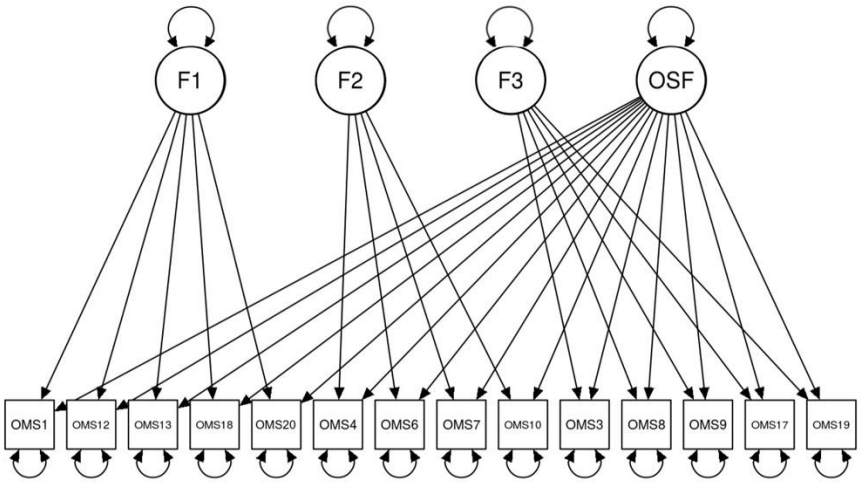
Modgill et al.'s (2014) three-factor structure for the OMS-HC-15.



Note. F1 = Attitude; F2 = Disclosure and Help-Seeking; F3 = Social Distance.

Figure 5

Öri et al.'s (2020) best-fitting model, bifactor structure, for the OMS-HC-14.



Note. F1 = Attitude; F2 = Disclosure and Help-Seeking; F3 = Social Distance; OSF = Overall Stigmatising Factor.

Inference Criteria. Model fit in the cited validation studies has been evaluated differently, as was outlined in Chapter 1. Chiba et al. (2016) used the goodness of fit index (GFI), adjusted goodness of fit index (AGFI), comparative fit index (CFI) and the Akaike's information criterion (AIC). In contrast, Öri et al. (2020) employed the root mean square error of approximation, the CFI and Tucker Lewis index (TLI). In the present study, a standard was set for uniformity purposes and that aligned with best practices (i.e., Hussey & Hughes, 2020). The goodness of fit was assessed with the following criteria from Hu and Bentler (1999): standardised root mean square residual (SRMR $\leq .09$), root mean square error of approximation (RMSEA $\leq .06$), Tucker Lewis index (TLI $\geq .95$), and comparative fit index (CFI $\geq .95$). With the definitions of Hussey and Hughes (2020), *Hypothesis 1a* will be supported if one or two but not all three metric

permutations (i.e., SRMR and RMSEA, SRMR and TLI, and SRMR and CFI) of Hu and Bentler's (1999) cutoff are satisfied. *Hypothesis 1b* will be supported if the two-factor structure has more metric permutations met than the unidimensional model. In contrast, *Hypothesis 2a* will be supported if all metric permutations are met. Further, *Hypothesis 2b* will be supported when the bifactor solution has more metric permutations met than the three-factor structure. *Hypotheses 2c* and *2d* will be considered supported if none of the metric permutations is met.

Although not considered as part of the inference criteria for assessing model fit, Chi-square values and their corresponding ρ values, normed chi-square and the 95% confidence intervals of the RMSEA will be calculated and reported to be aligned with the best practices in Hussey and Hughes (2020).

Model-based Reliability

The model-based reliability of the bifactor solution for the OMS-HC-14 was estimated. For this estimation, the ω_h, the explained common variance (ECV) and PUC were calculated. Reise et al. (2013) recommended using these indices to assess whether the multidimensionality of a measure is insufficient to disqualify the measure as unidimensional. The BifactorIndicesCalculator package (Dueber, 2017) in R was used for the model-based reliability analysis.

Inference Criteria. Using the cutoffs recommended by Reise et al. (2013), *Hypothesis 2e* will be supported if PUC < .8, ECV > .6, and ω_h > .7.

Internal Consistency

Internal consistency of the RAQ-7 and OMS-HC-15 was estimated to provide evidence of reliability of these two measures. McDonald's (1999) omega total (ω_t) was used as the primary

estimate of the reliability of the RAQ-7 and OMS-HC-15 total and subscale scores. However, the omega hierarchical (ω_h) and Cronbach's α were reported along with the ω_t. The selection of the primary and other estimates is in line with contemporary best practices (e.g. Dunn et al., 2014; Hussey & Hughes, 2020), as coefficient omega depends on a more sensible assumption about short items (items are not of tau equivalent structure but are congeneric). The 95% confidence intervals of the internal consistency reliability estimates were reported in line with Hussey and Hughes' (2020) recommendation. The present study did not replicate the inter-item correlation for the RAQ-7 done by Chiba et al. (2016). This was because the correlation between items on a short scale that assesses a complex construct cannot be high, which may cause the scale to have low internal consistency (Ziegler et al., 2014).

Inference Criteria. *Hypotheses 1d, 2f, 2g, 2h* and *2i* were based on Nunnally and Bernstein's (1994) cutoff, where values ≥ .7 signify good internal consistency.

Test-retest Reliability

Test-retest reliability of the RAQ-7 and OMS-HC was estimated to provide further evidence of reliability for the two measures. Unlike the two validation studies, Cohen's kappa statistic (1960) and the ICC were not used, given the goal to follow best practices as recommended in Hussey and Hughes (2020). The 95% confidence intervals of the test-retest reliability estimates were reported in line with Hussey and Hughes' (2020) recommendation.

Inference Criteria. Similar to internal consistency, *Hypotheses 1e, 2j* and *2k* were based on Nunnally and Bernstein's (1994) cutoff, where values ≥ .7 signify good test-retest reliability.

Shared Conceptual Theme

As with Ori et al. (2020), Spearman correlations were used to examine the monotonic relationships between the general and specific factors and within specific factors themselves. The psych package (version 2.1.9; Revelle, 2021) in R was used for this analysis and the internal consistency and test-retest reliability estimation.

Inference Criteria. For the intercorrelations between factors, cutoffs used by Öri et al. (2020) were employed whereby $r_s \leq .19$ are very weak, $.2 \leq r_s \leq .39$ are weak, $.4 \leq r_s \leq .59$ are moderate, $.6 \leq r_s \leq .79$ are strong, and $.8 \leq r_s \leq 1$ are very strong.

Chapter 3: Results

There were 320 participants who completed the survey. Of those, 34 did not work in the medical field, leaving 286 participants. Seven did not complete one to two items in the test, and one respondent missed out one item in the retest. However, no participant data were excluded based on missing data criteria (i.e., > 4). Of the 286 participants, all 19 invited took part in the retest. Table 3 provides participant demographic information. Nearly 75% of respondents were between 18 to 30 years old, 71% were female, and 63% were European.

Table 3

Participant Demographic Information

Variable	*N*	%
Age		
18-30	213	74.48
31-40	40	13.99
41 or older	33	11.54
Gender		
Male	82	28.67
Female	202	70.63
Non-binary/Third gender	2	0.70
Ethnicity		
African/African American/Black British	37	12.94
European	181	63.29
White/Sephardic Jew	33	11.54
Other[a]	35	12.24

Note. [a]Other includes several ethnicities (see Appendix C). Only the most common three were shown here. There were no Māori and Pacific Island participants. About half of the samples (53%) were from the UK.

The descriptive score data for the RAQ-7 and OMS-HC are shown in Table 4. The mean RAQ-7 total and subscale scores were skewed towards the highest possible scores, which indicates more positive attitudes towards recovery. On the other hand, the centres of the OMS-HC scores are leaning a bit towards the lowest possible scores, which signifies less stigmatising attitudes and behavioural intentions.

Table 4

Descriptive Statistics of the Total and Subscale RAQ-7 and OMS-HC Scores

Scale	*M*	*SD*	Min	Max	Possible Ranges
RAQ-7 Total	28.13	3.00	19	35	(7-35)
Recovery is possible and needs faith	14.59	2.41	6	20	(4-20)
Recovery is difficult and differs among people	13.53	1.22	10	15	(3-15)
OMS-HC-15 Total	31.86	7.80	15	62	(15-75)
Attitude	12.79	3.73	6	25	(6-30)
Disclosure and Help-seeking	10.31	3.16	4	18	(4-20)
Social Distance	8.76	2.83	5	21	(5-25)
OMS-HC-14 Total	29.67	7.26	14	57	(14-70)
Attitude (without item 14)	10.60	3.17	5	20	(5-25)

Note. Both the RAQ-7 and OMS-HC have no normative data. However, higher RAQ-7 scores indicate more positive beliefs about recovery, whereas lower OMS-HC scores denote less stigmatising attitudes.

Data were checked for multivariate normality before conducting the CFA. The results for Mardia's (1970) skewness and kurtosis test using the psych package (version 2.1.9; Revelle, 2021) for the RAQ-7 data were 373.28 and 6.49, respectively, and for the OMS-HC were 2,057.88 and 21.32, respectively. These results showed that the RAQ-7 and OMS-HC data were not multivariate normal. The results did not change the planned subsequent methodologies to demonstrate transparency and avoid spurious analyses, which can lead to equivocal findings, consistent with the recommendations in Flake and Fried (2020) and Hussey and Hughes (2020). The analyses in the current study were completed using untransformed raw scores, given that the distribution of the RAQ-7 (e.g., Chiba et al., 2016) and OMS-HC (e.g., Öri et al., 2020) scores in past researchers did not follow a normal distribution and/or was subsequently analysed nonparametrically.

Recovery Attitudes Questionnaire

Validity

The unidimensional and two-factor structures were tested to arrive at the more appropriate model for the RAQ-7. However, both models demonstrated poor fit. As shown in Table 5, only the SRMR values met the cutoff (i.e., SRMR ≤ .09, RMSEA ≤ .06, CFI ≥ .95, and TLI ≥ .95). Therefore, *Hypotheses 1a* and *1b* were not supported. Although not part of the preregistered inference criteria, fit indices of the two-factor structure were closer to the set cutoffs than the unidimensional model.

Table 5

Confirmatory Factor Analyses on the RAQ-7

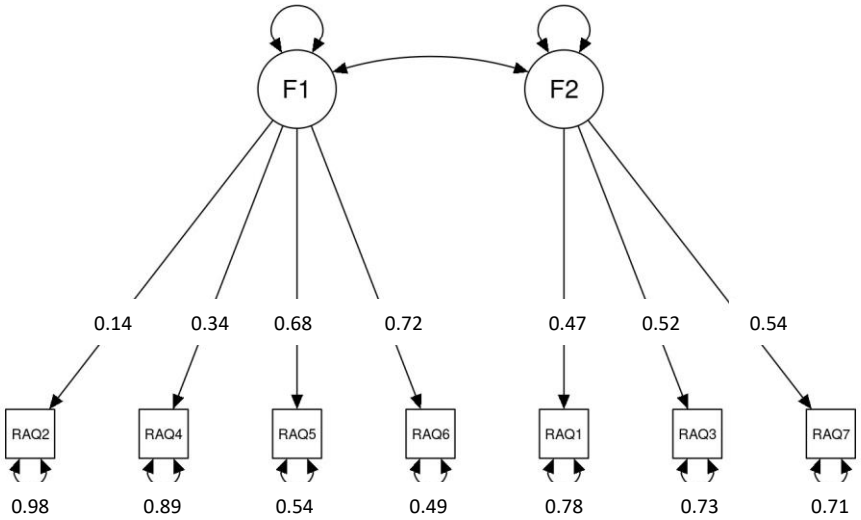
Model	χ^{2a}	χ^2/df	*df*	CFI	TLI	RMSEA [95% CI]	SRMR
Two-factor	34.205	2.93	12	0.753	0.761	0.082 [0.052, 0.114]	0.066
One-factor	58.700	5.24	11	0.479	0.474	0.122 [0.095, 0.151]	0.086

Note. χ^2 = Chi-square statistic, χ^2/df = Chi-square to degrees of freedom ratio, *df* = degrees of freedom, CFI = comparative fit index, TLI = Tucker-Lewis fit index, RMSEA = root mean square error of approximation, CI = confidence interval, and SRMR = standardised root mean squared residual. [a]For all χ^2 tests, $\rho \leq .001$.

The factor loading of item 2, "To recover requires faith", in the two-factor structure was very small (.14; see Figure 6), supporting *Hypothesis 1c* that the factor loading of item 2 will not exceed .3. Also, item 4, "Recovery can occur even if symptoms of mental illness are present", had a relatively small loading (.34).

Figure 6

Path diagram of the two-factor structure on the RAQ-7, showing standardised coefficients.



Note. F1 = Recovery is possible and needs faith*;* F2= Recovery is difficult and differs among people.

Reliability

The lower bounds of the 95% confidence intervals of the RAQ-7 total and factor scores' internal consistency (ω_t) and test-retest reliability (r) coefficients were below .7 (see Table 6). The hypotheses on replicating unsatisfactory internal consistency (*Hypothesis 1d)* and temporal consistency (*Hypothesis 1e)* were supported. The internal consistency reliability estimates range from moderate to strong, and the test-retest reliability estimates were very weak negative values.

Table 6

Reliability Estimation on the RAQ-7

Scale	Number of Items	α	ω_t	ω_h	r^a
Total	7	.58 [.23, .86]	.68 [.60, .98]	.30 [-.09, .78]	-.14 [-.56, .33]
Factor 1	4	.37 [.24, .94]	.51 [.49, .99]	.27 [.10, .87]	-.13 [-.55, .34]
Factor 2	3	.51 [.35, .95]	.52 [.45, .97]	.50 [-.21, .94]	.14 [-.33, .56]

Note. α= Cronbach's alpha, ω_t = McDonald's omega total, ω_h = McDonald's omega hierarchical, *r* = Pearson's correlation coefficient. For brevity, factor names were excluded. Values in square brackets were the lower and upper bounds of the 95% confidence intervals, respectively. [a]Test-retest interval was two weeks.

Opening Minds Scale for Health Care Providers

Validity

The data were fitted to the three models of the OMS-HC. As summarised in Table 7, the three-factor solution demonstrated a mixed fit because one (i.e., SRMR ≤ .09 + RMSEA ≤ .06) of the three metric permutations in the inference criteria was met. In contrast, the unidimensional model showed a poor fit (i.e., only SRMR met the cutoff). Thus, *Hypothesis 2d* was supported, given that the unidimensional model did not meet any of the three metric permutations. However, *Hypothesis 2c* was not supported, as the three-factor model did not exhibit a poor fit.

Table 7

Confirmatory Factor Analyses on the OMS-HC-15

Model	χ^{2a}	χ^2/df	*df*	CFI	TLI	RMSEA [95% CI]	SRMR
Bifactor (14 items)	-	-	-	-	-	-	-
Three-factor (15 items)	77.950	1.84	42	0.754	0.897	.054 [.042, .066]	0.057
One-factor (15 items)	126.104	3.01	42	0.418	0.755	.084 [.073, .095]	0.080

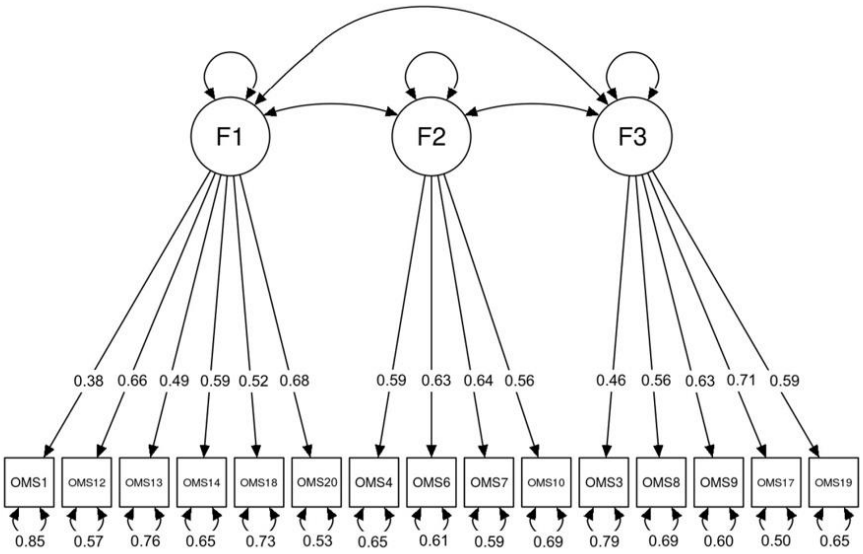
Note. χ^2 = Chi-square statistic, χ^2/df = Chi-square to degrees of freedom ratio, *df* = degrees of freedom, CFI = comparative fit index, TLI = Tucker-Lewis fit index, RMSEA = root mean square error of approximation, CI = confidence interval, and SRMR = standardised root mean squared residual. [a]For all χ^2 tests, $\rho \leq .001$.

On the other hand, a solution was not found when the current study data were fitted to the bifactor model. A rerun of the analysis with the current study data using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm instead of the default optimisation method was done, consistent with a recommendation when lavaan warns of an unattainable solution (Lin, 2021). The BFGS employs more iterations than the default method but could obtain similar results (with slight rounding differences) as the default method (Lin, 2021). This rerun rendered negative latent variable variances with a specific factor (i.e. Factor 1 – *Attitude*). A further inspection was done by performing a simulation with synthetic data from the lavaan package (Rosseel, 2012), to ensure that the unattainable solution was not due to an incorrect R code. This simulation resulted in the calculation of fit indices. Since fit measures were not available, *Hypotheses 2a* and *2b* were not supported.

In terms of the factor loading, item 1 – "I am more comfortable helping a person who has a physical illness than I am helping a person who has a mental illness" – loaded the weakest (i.e., .38) on the Attitude factor in the three-factor structure (see Figure 7).

Figure 7

Path diagram of the three-factor structure on the OMS-HC-15, showing standardised coefficients.



Note. F1= Attitude; F2 = Disclosure Help-Seeking; F3 = Social Distance.

Reliability

All the lower bounds of the 95% confidence intervals for the total and subscale scores on the OMS-HC except for one (i.e., Factor 1 – *Attitude* – without item 14) were above .7 (see Table 8). This meant that *Hypotheses 2f, 2g* and *2h* (and not *2i*) were supported, as the lower bounds of

e

internal consistency estimates in almost all OMS-HC scales were strong. On the other hand, contrary to *Hypotheses 2j* and *2k*, the total and subscale scores' test-retest reliability lower bounds were negative values. A strong but negative correlation was seen for factor 2 – *Disclosure and Help-seeking* – with -.75 lower bound of the 95% confidence interval.

Table 8

Reliability Estimation on the OMS-HC-15

Scale	Number of Items	*a*	ω_t	ω_h	r^a
OMS-HC-15	15	.84 [.73, .92]	.86 [.86, .98]	.72 [.20, .74]	-.24 [-.63, .24]
Factor 1	6	.73 [.41, .85]	.77 [.76, .97]	.60 [.33, .77]	-.20 [-.60, .28]
Factor 2	5	.70 [.36, .86]	.76 [.75, .94]	.62 [.18, .89]	-.46 [-.75, -]
Factor 3	4	.74 [.64, .91]	.77 [.71, .97]	.67 [.57, .83]	.21 [-.27, .61]
OMS-HC-14	14	.82 [.75, .90]	.85 [.84, .97]	.45 [.23, .79]	-.29 [-.66, .19]
Factor 1 (no item 14)	5	.69 [.01, .75]	.73 [.61, .90]	.63 [.25, .78]	-.29 [-.66, .19]

Note. α = Cronbach's alpha, ω_t = McDonald's omega total, ω_h = McDonald's omega hierarchical, *r* = Pearson's correlation coefficient. For brevity, factor names were excluded. Values in square brackets were the lower and upper bounds of the 95% confidence intervals, respectively. [a]Test-retest interval was two weeks.

As the bifactor model did not converge (see results on Validity), no standardised coefficients were produced for the model-based reliability computation. When further iterations were made, there were negative latent variable variances for factor 1 – *Attitude*, and, hence, its

model-based reliability estimates were unavailable. Nevertheless, Table 9 shows the obtained values. *Hypothesis 2e* was not supported because complete model-based reliability estimates were not available.

Table 9

Model-based Reliability Estimation on the Bifactor Solution for the OMS-HC

PUC	ECV			
	Overall Stigmatising Factor	Attitude	Disclosure & Help-seeking	Social Distance
0.82	0.65	-	0.66	0.42

Note. PUC = percent of uncontaminated correlations; ECV = explained common variance

Shared Conceptual Theme

The associations between the three OMS-HC subscale scores (without item 14) and total score were examined to obtain evidence for the "shared conceptual theme" (Modgill et al., 2014, p. 8). Table 10 shows that the correlation coefficients for the relationships between the general factor (i.e., "overall stigmatising attitude"; Őri et al., 2020, p. 5) and the specific factors exceeded the set cutoff (i.e., $r_s > .59$), supporting *Hypothesis 2l.* On the other hand, the relationship between the specific factors ranged from weak to moderate. *Hypothesis 2m* was not supported since not all of the correlations between these specific factors were $r_s \leq .39$.

Table 10

Correlation Coefficients between the OMS-HC Total and Subscale Scores

Relationship	r_s
General Factor and Factor 1	.83 [.79, .87]
General Factor and Factor 2	.75 [.69, .80]
General Factor and Factor 3	.76 [.71, .81]
Factors 1 and 2	.45 [.34, .54]
Factors 1 and 3	.55 [.46, .62]
Factors 2 and 3	.32 [.22, .42]

Note. r_s = Spearman's correlation coefficient. For brevity, factor names were excluded. Values in square brackets were the lower and upper bounds of the 95% confidence intervals, respectively.

Chapter 4: Discussion

The current study evaluated two measures of attitudes regarding personal recovery and mental distress stigma, the RAQ-7 (Borkin et al., 2000) and the OMS-HC (Kassam et al., 2012; Modgill et al., 2014), respectively. Open Science principles were used to address the methodological limitations of previous validations of these two measures. CFAs, with multiple measures of model fit, were conducted to examine whether the proposed factor structures in original and replication studies could be found in a sample working in the medical sector. Additionally, the internal consistency and test-retest reliability of the two measures were investigated.

Recovery Attitudes Questionnaire

The validation of the RAQ-7 yielded equivalent results to the study undertaken by Chiba et al. (2016) in Japan. Thus, the hypotheses on unsatisfactory reliability were supported but not for the equivocal fit to the known and competing structures. The study confirms concerns regarding the validity and reliability of the seven-item scale. A more detailed discussion follows.

Validity

The data in the current study poorly fitted the two-factor structure of Borkin et al. (2000), replicating the results of the Japanese study. Only the absolute fit indices (i.e., SRMR in the current study; GFI and AGFI in the Japanese research) reached the benchmarks. Using a single fit index has proven ineffective in establishing structural validity (Hu & Bentler, 1999; Hussey & Hughes, 2020; Vandenberg & Lance, 2000), so the findings do not provide evidence of the validity of the known structure for the RAQ-7. Similarly, the unidimensional solution met only

one metric, despite earlier investigation concluding that it is an adequate alternative to explain the structure of the RAQ-7 (e.g., Hungerford et al., 2015; Wilrycx et al., 2012). Items on the scale may be more relevant to some medical workers than others, considering that mental health recovery attitudes vary across age, training, clinical background, occupation, and contact with people with lived experience of mental distress (Luigi et al., 2020). A portion of the medical workers could have a more favourable attitude towards recovery than others, and such difference could probably be due to differences in professional experience with people with mental distress. Differences in the appreciation of items on a measure may reflect that the construct holds a different meaning for the population and could result in the non-confirmation of a measurement model (DiStefano & Hess, 2005). More evidence is needed to support the construct validity of the RAQ-7, including testing measurement invariance for differences in the construct(s) interpretation across demographics, which was not performed in the current study. Testing for measurement invariance could help isolate what specific group with, say, similar gender, culture or development could share a two-factor structure understanding of recovery. Also, given the multivariate non-normality of the RAQ-7 scores, it is incumbent that future studies will have to use nonparametric tests, such as CFAs with robust estimators that were used in the current study.

Item 2, "To recover requires faith", did not fit well (i.e., .14) within the two-factor model, which aligned with previous research (Chiba et al., 2016; Hungerford et al., 2015; Wilrycx et al., 2012). Also, this item loaded poorly (i.e., .27) within the unidimensional model examined in the current study. Although the RAQ-7 used in the current study was presented verbatim as it was developed, medical sector workers may hold different views of "faith". They may not see Item 2 in the same conceptual way as the American samples in the development of the RAQ-7. The idea of "faith" can be influenced by culture (Brown, 2008), meaning-making (Dyess, 2011) and the

sciences (Koenig, 2004). For those "faithful" to the sciences, evidence-based treatments (Newlin et al., 2012; Salsman et al., 2012), hierarchies of authority (Vuckovich & Artinian, 2005), and acceptance of biomedical perspectives in mainstream health services (Lakeman, 2013) may add nuances to their interpretation of "faith". Thus, in a non-American context, such differences in the idea of "faith" render a question on the fitness of associating Item 2, a very broad four-word stem, with the construction of recovery. In other words, is "faith" a core concept of recovery to begin with? Further, it is unclear how medical workers or health care professionals in a given secular or non-secular culture demonstrate positive attitudes towards recovery in practice. There needs to be a contextualised definition of "faith" and other fundamental concepts in the RAQ-7 to be valid across cultures.

Reliability

Similar to Chiba et al. (2016), the internal consistency reliability of the RAQ-7 was unsatisfactory. Although the estimates were moderate to strong (i.e., ω_t = .52 - .68), comparable to those from previous studies (Jaegar et al., 2013; Rabenschlag et al., 2014; Wilrycx et al., 2012), the lower bounds of the 95% confidence interval were less than the .7 cutoff. This internal consistency was contrary to studies that used the English version (Hardiman & Hodges, 2008; Hungerford et al., 2015; Salgado et al., 2010), where Cronbach's α coefficients ranged from .72 to .74. According to the RAQ-7 developers, poor estimates are likely due to a limited number of test items. The current research supports this claim, as evidenced by the dwindling coefficients and lower bounds for each scale and subscale as the number of items decreases. However, increasing the number of items may be at odds with the original plan of the authors of the measure to create a brief measure of attitudes towards recovery. Hungerford et al. (2015) suggested an alternative to bring the estimates above the recommended standard, by removing

item 2 ("To recover requires faith") and item 3 ("Stigma associated with mental illness can slow down the recovery process"). Nevertheless, item deletion decisions may need to be made based on rigorous statistical and qualitative criteria (Wieland et al., 2017), considering that a subscale requires a minimum of three items to be valid (Korlén et al., 2018; Rakov & Marcoulides, 2000; Spector, 1992). In other words, reducing the "*Recovery is difficult and differs among people*" subscale to two items by removing item 2 ("To recover requires faith.") might negatively affect its factorial validity.

The test-retest reliability estimates and lower bounds for two weeks of the RAQ-7 were very weak to moderate negative values. The findings suggest that the scores obtained from the measure may not be stable over time. Current values differ from previous findings but consistently demonstrate that the RAQ-7 does not have strong temporal stability (Borkin et al., 2000; Chiba et al., 2016). As test-retest reliability was not established, assertions that the RAQ-7 could be used to evaluate changes over time (e.g., Wilrycx et al., 2012) could not be made. To adequately explain the test-retest reliability of the measure, intervening factors (e.g., interventions, practice effect and response shift) between test administrations must be taken into account (Cohen & Swerdlik, 2009). Firstly, it is essential to note that the present study, like previous validations that demonstrated evidence of temporal stability for the RAQ-7, had a sample that was not subjected to intervention between measurement occasions. Since no intervention occurred and it was unknown whether samples had been exposed to interventions outside of the study, the potential practice effect could be a possible explanation. Nevertheless, it did not appear that the prior testing had rehearsed the samples for better performance on the RAQ-7. The recovery attitudes were not necessarily improved, as nearly half of the retest scores were lower (i.e., more negative) than the first test scores. Also, the contention that the retest

findings could be due to response shift, change in the individual's evaluation of the construct that is remedied by a shorter retest interval (Schwartz & Rapkin, 2004; Schwartz & Sprangers, 2000) is open to question. Like Chiba et al. (2016), the current time frame, shorter than the developer's 19 days, generated similarly unacceptable results. Given that the potential significant challenges may not have affected the retest scores, future studies examining this form of reliability may consider a follow-up session, similar to the suggestion by Polit (2014), which is an opportunity for test users to ask the test takers for the reason(s) of their response regularities and irregularities. Lastly, performing a test of dependability, another form of test-retest reliability estimation, where a retest is taken within an hour but no later than 24 hours after taking the initial test (Revelle & Condon, 2018), would allow for more comparison of the best possible interval between measurements that enhances the accuracy of reliability estimation.

Using the RAQ-7 may pose challenges in assessing attitudes towards recovery, as it is neither a measure of precision nor stability. To be effective, it must at least possess reliable psychometric properties (e.g., internal consistency, test-retest reliability; Scheyett et al., 2013). In addition, the measure was found to have no acceptable structural validity. These findings point out the inherent complexity of the construct of recovery, which is difficult to measure appropriately (Law et al., 2012; van Weeghel et al., 2019; Stevens Manser et al., 2018), especially when there is no single definition of recovery. A measure with seven items may all the more not effectively represent the different facets of recovery. Thus, the RAQ-7 may need to be re-examined to be a good measure for baseline recovery attitudes in a single administration or shifts in perspectives over time due to any intervention. Otherwise, test users will have to utilise a valid and reliable alternative to it.

Opening Minds Scale for Health Care Providers

The current study was a partial replication of Őri et al. (2020; Hungarian study). Three models were examined to determine the best structure to explain the construct behind the OMS-HC, internal and temporal reliabilities were estimated, and the interrelationships between the factors of the measure were tested. The findings did not confirm all model fits, reliability evidence and interrelationships found in the validation of Őri et al. (2020). Therefore, the only hypotheses supported in the study were that the data would possess satisfactory internal consistency and a poor fit to the unidimensional model.

Validity

Among the models tested, the correlated three-factor structure for the OMS-HC by Modgill et al. (2014) met the most metric standards (i.e., SRMR ≤ .09 + RMSEA ≤ .06). This mixed fit result contradicts the Hungarian study, where only the RMSEA exceeded the cutoff. Instead, the findings appear to be closer to those of Sapag et al. (2019; Chile), in which the model met the SRMR criterion and one of the other three metrics. The current findings were also similar to the Chilean study in sample heterogeneity (i.e., diverse public health care staff and providers). A potential explanation for the mixed fit is that the OMS-HC-15 could measure, on top of stereotypes, prejudice, and discrimination, a fourth stigma mechanism or construct under the stigmatised perspective, anticipated stigma (Fox et al. 2018). The fit indices still provided partial evidence that the OMS-HC-15 comprises three subscales - *Attitude, Disclosure and Help-seeking*, and *Social Distance*, notwithstanding the equivocation. Within the same three-factor model, item 1, "I am more comfortable helping a person who has a physical illness than I am helping a person who has a mental illness", had the weakest loading, though not below .3, among the indicators. A similar result was reported by Chang et al. (2017; Singapore). If further

explored, the three-factor structure can generate improved model fit indices, as was the case in the Singaporean study, where an EFA was conducted after eliminating item 1.

In contrast, the current data did not fit well with the unidimensional structure, with only the absolute fit index meeting the cutoff point. The Hungarian study, in which none of the fit indices met the benchmarks, and a study from North-western Europe and Australia (Happell et al., 2019) reported similar findings. Based on these findings, the overall OMS-HC-15 score is not a valid indicator of mental distress stigma attitudes. Researchers evaluating anti-stigma interventions and other test users should reassess using the total OMS-HC-15 score. Consideration should be given to supporting a unidimensional structure, or choosing alternative measures when using a single score is crucial to determining overall levels of stigma and subsequent intervention efficacy.

The bifactor solution proposed by Őri et al. (2020) was also examined. Items in this model were expected to load on a general factor ("overall stigmatising attitude"; p. 5) and a specific factor, one of the three in Modgill et al. (2014). However, fit indices and complete model-based reliability could not be obtained with the current data. Negative latent variable variances were derived from further iterations with an alternative optimisation method. Thus, the results raise the question of whether the general score and three orthogonal subscale scores can be used in whatever context is relevant. In other words, test users could not have the ability to indicate the OMS-HC is unidimensional enough to represent an overall stigmatising attitudes score or to report that the measure is multidimensional enough to represent a subscale score, say, Factor 3 - *Social Distance*, on its own. Several possible factors can be explained as the cause of this result, including model misspecification (e.g. Bollen, 1989; Boomsma & Hoogland, 2001; Rindskopf, 1984) and sampling variabilities (e.g., Anderson & Gerbing, 1988; Bollen, 1987). If

further explored, various methods could be employed to identify the source of the improper solution (Kolenikov & Bollen, 2012). As the bifactor model of the OMS-HC was constructed with a homogenous sample (i.e. practising psychiatrists), future studies exploring this structure may be able to replicate that the model is valid with similar sample composition to the Hungarian study.

Reliability

The OMS-HC-15 total and subscale scores demonstrated strong internal consistency, as evidenced by the estimates and lower bounds of 95% confidence interval exceeding the cutoff. The results aligned with Destrebecq et al. (2018) and exceeded the developers' findings. Further, this is the first time the factor 2 subscale - *Disclosure and Help-seeking*, which has the fewest items and often has the lowest reliability estimates among the subscales, registered above .7. Removing an item from the measure, as in the OMS-HC-14, was associated with a strong total score but weak modified factor 1 – *Attitude* (without item 14) score internal consistency. Nevertheless, given the factorial validity findings, using highly reliable total scores might need rethinking.

The test-retest reliabilities of the OMS-HC total and subscale scores were unsatisfactory, ranging from weak to strong negative lower bounds of the 95% confidence intervals. These negative stability coefficients might indicate events that have impacted participants. Similar to the retest findings of the RAQ-7, the fluctuation could not be attributed to a deliberate intervening effort to change the scores and practice effect that could have prompted the participants to think that the retest was a task that tested their ability to be consistent. In addition, the two-week interval could not be blamed for the negative coefficients considering that such is typical for retest research (Polit, 2014). Lastly, the results could not be attributed solely to the

number of participants, given the sample size planning before data collection. However, if resources are not limited (i.e., there is an ability to pay more Prolific Academic participants to partake in the study), there is no harm in conducting the retest with a larger sample. This test-retest reliability estimation has been one of the few, if not the first, with a two-week interval and, if not the second, with the OMS-HC-15. Destrebecq et al. (2018) examined the same reliability in one week but found strong estimates.

Shared Conceptual Theme

The notion of a "shared conceptual theme" in the OMS-HC, as noted by Modgill et al. (2014, p. 8) and Őri et al. (2020, p. 7), has some support in the current study, albeit weak to moderate (r_s = .32 - .55). The correlations between the subscale factors illustrate the interconnectedness of stigma's cognitive, affective, and behavioural mechanisms that represent a person's response to another in terms of devaluing their identity (Barney et al., 2010; Fox et al., 2018; Michaels et al., 2017). Therefore, an increase in factor 1 (*Attitude*) scores with items on stereotypes regarding people with mental distress, the cognitive mechanism, and on negative emotional reactions, the affective mechanism, may reflect an increase in factor 2 (*Disclosure and Help-Seeking*) scores on anticipated stigma and factor 3 (*Social Distance*) scores on intention to discriminate, the behavioural mechanism, and vice-versa. For example, medical workers may think (cognitive) people with lived experience of mental distress are 'unpredictable' or 'dangerous', as has been reported in prior stigma research (e.g., Serafini et al., 2011; Subramaniam et al., 2017; Yap et al., 2014). These beliefs may lead to fear (affective) as the common form of prejudice towards people with mental distress (Corrigan, 2005), including the fear that others, such as the wider public, may perceive them (i.e., medical workers) as 'unpredictable' or 'dangerous' as well. Consequently, medical workers may expect that stigma

might apply to them (i.e., making them the stigmatised) and, thus, reduce the frequency of discussing personal lived experiences of mental distress with colleagues (behavioural), which has been an identified outcome of self-stigma (Corrigan & Rao, 2012).

Overall, the OMS-HC has some good psychometric properties. The high internal consistency and some support for the three-factor structure and "shared conceptual theme" (Modgill et al., 2014, p. 8; Őri et al., 2020, p. 7) may be helpful for future research into mental distress stigma. In addition, a reliability profile of high internal consistency and very low temporal consistency, as for the OMS-HC, could mean that the instrument measures a state or mood construct that varies across time (Widaman et al., 2011; Steyer et al., 2012). If such is the case, it can be questioned whether the OMS-HC is an outcome measure, consistent with how most health measures, like the quality of life measures, tend to assess state-like attributes (Mokkink et al., 2010). However, there is no clear indication in previous development and validation studies as to whether attitude, as measured in the OMS-HC, is a latent trait or state. Hence, replicating this validation with a more robust method can aid in identifying whether subscale scores remain more valid and reliable indicators of construct change over time.

Ongoing Issues

Since the factorial validity and reliability of the RAQ-7 and OMS-HC in this study were mixed or poor, questions may be raised regarding the process of previous validations in concluding that the measures were psychometrically valid. Although the tests performed in this study may be regarded as more robust, there is the possibility that prior studies may have underreported specific evidence of a psychometric property by using metrics that are more widely and conveniently available and thereby falsely claim structural validity, as has been determined by previous research (e.g., Crutzen & Peters, 2017; Flake et al., 2017) and discussed

in Chapter 1. Previous validators of the two measures could have focused their efforts and resources on specific aspects of structural validity rather than other more critical psychometric properties, consistent with the argument of Hussey and Hughes (2020). If such is the case, the importance of understanding the profile of the psychometric properties of a measure and matching it with the needs of the setting where the measures are to be used could have also been overlooked - a task that Ziegler et al. (2014) emphasised to be more critical to asking whether or not a measure is psychometrically valid or not. For instance, before investigating a given psychometric property, a test user or evaluator must optimise internal consistency if a measure is intended for individual assessments, whereas a high test-retest reliability must be given premium if a measure is used for group comparisons (Kruyen et al., 2012). Lastly, poor or mixed validity and reliability could have resulted from an inadequate definition of the scales' constructs, another practice issue that researchers recognise as particularly relevant in designing and evaluating measures (Ziegler et al., 2014). The lack of appropriate definitions as a potential cause for equivocal outcomes would mean that there are problems not only with the structural phase but also with the substantial phase of validation (Flake et al., 2017) for the RAQ-7 and the OMS-HC, as discussed in the Chapter 1. In any case, disagreements between findings and the increasing number of possible explanations for those findings are likely to be affected by the diversity of underreported measurement practices used in research.

Another recurrent theme underlying the equivocal findings is the inherent subjectivity of personal recovery and mental distress stigma. It is challenging to develop a consensus definition on complex and multidimensional constructs such as recovery and stigma, including whether they involve "faith". These constructs are so complex that there are several ways to define them (e.g., recovery vision of Anthony [1991]; CHIME of Leamy et al. [2011]; stigma

conceptualization of Link and Phelan [2001]; stigma framework of Thornicroft [2006]). Yet, even these broad frameworks that attempt to comprehensively explain these constructs receive criticisms that they are still not all-encompassing (e.g., CHIME-D of Stuart et al. [2017] in response to CHIME; mental illness stigma framework of Fox et al. [2018]). Furthermore, the complexity of these constructs has not only prompted different researchers to come up with their measures but also come up with measures and items that assess multiple factors within a subscale, consistent with the evidence on how specific measures conflate one stigma construct with another (Fox et al., 2018) and earlier discussed equivocal findings for the known three-factor structure for the OMS-HC. In other words, measures such as the RAQ-7 and the OMS-HC could not optimally capture the construct (or constructs) under study because they may have resulted from the inadvertent measurement of overlapping but earlier underappreciated factors, which, in turn, stem from the inherent complexity of the construct. The difficulty of teasing out distinct factors within a measure because they may be closely related to other distinct factors could contribute to poor or mixed validity findings (Hussey & Hughes, 2020). In light of the above, it is pertinent to demonstrate a degree of tolerance for how these constructs may be interpreted. Through the integration of different components of an inclusive exchange process, such as the recognition of ambiguity and the development of a mutual understanding between participants (Seikkula et al., 2006), the curiosity to gain an in-depth explanation of a finding (Polit, 2014), and openness to disagreements (Hasson-Ohayon et al., 2017), a more encompassing meaning may be found. Implications of this difference in meaning are further discussed in the proceeding section.

Strengths, Limitations and Future Directions

The earlier discussion on the complexity of the conceptual underpinnings of the RAQ-7 and OMS-HC is vital in implementing these measures in practice. The ability of medical workers or health service providers to provide care is determined by their local context of personal recovery and mental distress stigma. Whether one measure covers all local contexts is not clear. More specifically, whether a measure with few items could comprehensively capture a facet of a bigger construct is not certain. Besides the multidimensional constructs at the front and centre of the two measures, neither developers of the two measures adequately addressed how an "attitude" is defined or at what level its influence is to be measured. Further, this concern is compounded because "attitude" fluctuates between separate measurement occasions (Polit, 2014; Widaman et al., 2011), as evidenced by the retest findings. However, it is unclear whether maturation brought about by professional practice and clinical experience or shifting values in the wider general public determine conceptualisation shifts (Ellison et al., 2018). Therefore, a more significant body of research is needed to determine the utility of the measures and inform decisions regarding when and how often measures of attitudes towards personal recovery and mental distress stigma should be employed.

On the other hand, despite the highly individualised nature of personal recovery, as signalled in the association of "attitudes" in the RAQ-7 with a "sense of self" by Borkin et al. (2000, p. 96), it is now more important than ever to pay attention to the conceptualisations of recovery in non-Western groups and indigenous people. Spirituality (De Ruysscher et al., 2017; Slade et al., 2012; Wood & Alsawy, 2018) and culture (Ellison et al., 2018) affect optimism and mental health positively but are not inherently woven into the predominantly secular construction of personal recovery by Western, middle-class professionals (Slade et al.,

2012). In addition, health literacy research continues to be lacking in locations outside of Western countries (Wei et al., 2015). Despite recruiting from an international pool, most of the participants in the current study were European. Further, the crowdsourcing platform from which the participants were recruited had worked predominantly based in the UK or the US. Therefore, the current findings may not apply to non-European medical workers and may not add value to the non-Western health research discussion. Future studies should incorporate a more diverse sample in development and validation studies and shift the focus of confirmatory research from individualistic preferences to other philosophical or theological worldviews that could explain personal recovery. There is a possibility that recovery measures based on spirituality and culture may be as effective, if not better, than more comprehensive recovery instruments or tests.

The importance of expanding on non-Western conceptualisations cannot be overstated, not only when it comes to recovery but also in stigma research. Improving the health and well-being of people with lived experience of mental distress requires understanding differences in perceptions and experiences of mental distress stigma. People of colour, who are more likely to face inadequate mental health care needs (Wang et al., 2007), hold stigmatising views about mental distress that are explained by their differences in socialisation and values (Abdullah & Brown, 2011; Rao et al., 2007). The cultural values that affect functioning in social environments may lead to more specific stigma measures than generic ones (Yang et al., 2007, 2014).

Using a sample of online respondents working in the medical sector has strengths and weaknesses. Researching with samples in the medical sector supports the call to expand personal recovery works to include the broader community and not just the people with lived experience of mental distress (Tew et al., 2012; Topor et al., 2011), considering that the promotion of recovery-oriented practices to combat mental distress stigma happens in an interpersonal and

social context (Chester et al., 2016). However, validating measures with medical workers who may or may not have contact with people with mental distress may mean that research on subjective interpretation of recovery concepts may or may not align with the definitions of practitioners who provide direct care to people with mental distress. While studies with specific samples may need more time and resources, it is still critical to research health care professionals with exposure to people with lived experience of mental distress for the understanding of recovery to be advanced, given the evidence that mental health professionals are a significant source of negative and stigmatising attitudes (Schulze, 2007). Measures that are not operationalised to systems comprised of mental health professionals that provide mental health care may render challenges in benchmarking what qualifies as recovery-centred service (Law et al., 2012). Further, if services are not implemented and measured in a more focused way, recovery orientation which has become the focus of mental health policies will not progress (Slade et al., 2012).

Validating measures with online samples seems relevant when considering the increasing dependence on self-reports of online respondents (Sassenberg & Ditrich, 2019). However, it is acknowledged that not all targeted populations of personal recovery and anti-stigma programmes that employ measures such as the RAQ-7 and OMS-HC, including professionals with exposure to service users, could be represented in one source (i.e., online crowdsourcing platform). Further, the sample characteristics were not maximised, and some other pertinent demographic questions were not obtained (e.g., Prolific Academic workers were not asked to identify their medical sector occupations). Future studies may need to replicate the analysis of the socio-demographic information by initial validation studies for both the RAQ-7 and OMS-HC, where additional information regarding the potential mental health diagnosis, treatment background,

and contact with people with lived experience of mental distress of the mental health professional is obtained. This analysis could have helped rule out or explain any sampling fluctuations for negative variances or Heywood cases and differentiate attitudes between groups. If analyses of attitudes are stratified by groups, known and competing factor structures may be validated, as evidenced by Modgill et al. (2014), who found that the OMS-HC-15 items loaded well onto three factors in different professional groups except for social workers.

The use and reporting of robust methods are a strength of this study. This study would be one of the few, if not the first, to apply model fit decision-making, preregister research protocols, and share supplementary materials (i.e., anonymised data and R codes) for both measures studied. In addition, the use of robust maximum likelihood estimation, which was not done in the past for the validation of the RAQ-7, was appropriate, given the scores on both measures in the current study were multivariate non-normal. Additionally, this study provided further information on the feasibility of a bifactor solution for the OMS-HC. However, global structural validity as advocated by Hussey and Hughes (2020) cannot be determined since measurement invariance and tests of dependability were not conducted. There have been few validations on personal recovery and stigma measures, and not all of these studies employed rigour and transparency. Further studies should emphasise the importance of validating existing measures over developing new ones and validating existing measures with well-deliberated and open methodologies.

Conclusion

In the current study, neither the factorial validity nor reliability of the RAQ-7 was conclusively demonstrated, whereas there was equivocal support for the known structure, shared conceptual theme, and reliability of the OMS-HC-15. In addition, the complexity of the

constructs at the core of these measures has major implications for accurate measurement. Finally, the process and outcome of validating these two measures highlight the importance of combining a robust and open measurement approach with the continued effort to gather and extend validity and reliability evidence to consolidate information about personal recovery and mental distress stigma theories into practice.

References

Abdullah, T., & Brown, T. L. (2011). Mental illness stigma and ethnocultural beliefs, values, and norms: An integrative review. *Clinical Psychology Review, 31*, 934–948. http://dx.doi.org/10.1016/j.cpr.2011.05 .003

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Andresen, R., Oades, L., & Caputi, P. (2003). The experience of recovery from schizophrenia: towards an empirically validated stage model. *The Australian and New Zealand Journal of Psychiatry, 37*(5), 586–594. https://doi.org/10.1046/j.1440-1614.2003.01234.x

Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal, 16*(4), 11–23. https://doi.org/10.1037/h0095655

Anthony, W. A., Cohen, M., & Farkas, M. (1999). The future of psychiatric rehabilitation. *International Journal of Mental Health, 28*(1), 48–68. https://doi.org/10.1080/00207411.1999.11449446

Arboleda-Flórez, J., & Stuart, H. (2012). From sin to science: Fighting the stigmatization of mental illnesses. *The Canadian Journal of Psychiatry, 57*(8), 457–463. https://doi.org/10.1177/070674371205700803

Arifin, W. N. (2018). A web-based sample size calculator for reliability studies. *Education in Medicine Journal*, *10*(3), 67-76. https://doi.org/10.21315/eimj2018.10.3.8

Asendorpf, J. B., Conner, M., de Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2016). Recommendations for increasing replicability in psychology. In A. E. Kazdin (Ed.), *Methodological Issues and Strategies In Clinical Research* (pp. 607–622). American Psychological Association. https://doi.org/10.1037/14805-038

Bambra, C., & Pope, D. (2007). What are the effects of anti-discriminatory legislation on socioeconomic inequalities in the employment consequences of ill health and disability? *Journal of Epidemiology & Community Health*, *61*(5), 421-426. http://dx.doi.org/10.1136/jech.2006.052662

Barney, L. J., Griffiths, K. M., Christensen, H., & Jorm, A. F. (2010). The Self-Stigma of Depression Scale (SSDS): Development and psychometric evaluation of a new instrument. *International Journal of Methods in Psychiatric Research*, *19*(4), 243–254. https://doi.org/10.1002/mpr.325

Barry, A. E., Chaney, B., Piazza-Gardner, A. K., Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior*, *41*, 12–18. https://doi.org/10.1177/1090198113483139

Bauer, D. J., & Curran, P. J. (2016). The discrepancy between measurement and modeling in longitudinal data analysis. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications* (pp. 3–38). IAP Information Age Publishing.

Bellack A. S. (2006). Scientific and consumer models of recovery in schizophrenia: Concordance, contrasts, and implications. *Schizophrenia Bulletin*, *32*(3), 432–442. https://doi.org/10.1093/schbul/sbj044

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78-117. https://doi.org/10.1177/0049124187016001004

Beresford, P. (2020). 'Mad', Mad studies and advancing inclusive resistance. *Disability & Society*, *35*(8), 1337-1342. https://doi.org/10.1080/09687599.2019.1692168

Berry, C., Gerry, L., Hayward, M., & Chandler, R. (2010). Expectations and illusions: A position paper on the relationship between mental health practitioners and social exclusion. *Journal of Psychiatric and Mental Health Nursing*, *17*(5), 411–421. https://doi.org/10.1111/j.1365-2850.2009.01538.x

Bird, V., Leamy, M., Tew, J., Le Boutillier, C., Williams, J., & Slade, M. (2014). Fit for purpose? Validation of a conceptual framework for personal recovery with current mental health consumers. *The Australian and New Zealand Journal of Psychiatry*, *48*(7), 644–653. https://doi.org/10.1177/0004867413520046

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and

behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bollen, K. A. (1987). Outliers and Improper Solutions: A Confirmatory Factor Analysis Example. *Sociological Methods & Research*, *15*(4), 375–384. https://doi.org/10.1177/0049124187015004002

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* John Wiley & Sons. https://doi.org/10.1002/9781118619179

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. *Structural Equation Models: Present and Future. A Festschrift in honor of Karl Jöreskog*, *2*(3), 139-168.

Borkin, J. R., Steffen, J. J., Ensfield, L. B., Krzton, K., Wishnick, H., Wilder, K., & Yangarber, N. (2000). Recovery Attitudes Questionnaire: Development and evaluation. *Psychiatric Rehabilitation Journal*, *24*(2), 95–102. https://doi.org/10.1037/h0095112

Borschmann, R., Greenberg, N., Jones, N., & Henderson, R. C. (2014). Campaigns to reduce mental illness stigma in Europe: A scoping review. *Die Psychiatrie*, *11*(01), 43-50.

Bos, A. E., Pryor, J. B., Reeder, G. D., & Stutterheim, S. E. (2013). Stigma: Advances in theory and research. *Basic and Applied Social Psychology, 35*, 1–9. http://dx.doi.org/10.1080/01973533.2012.746147

Boyd, J. E., Adler, E. P., Otilingam, P. G., & Peters, T. (2014). Internalized Stigma of Mental Illness (ISMI) scale: A multinational review. *Comprehensive Psychiatry, 55,* 221–231. http://dx.doi.org/10.1016/j .comppsych.2013.06.005

Brewer, M. B. (2007). The social psychology of intergroup relations: Social categorisation, ingroup bias, and outgroup prejudice. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology: Handbook of Basic Principles* (2nd ed., pp. 695–715). New York, NY: Guilford Press.

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, 10,* 80–91. http://dx.doi.org/10.1186/1472-6963-10-80

Brown, L. S. (2008). *Cultural Competence in Trauma Therapy: Beyond the Flashback.* American Psychological Association. https://doi.org/10.1037/11752-000

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (Second edition). New York; London: The Guilford Press.

Burgess, P., Pirkis, J., Coombs, T., Rosen, A., & Australian, A. J. (2010). *Australian Mental Health Outcomes and Classification Network 'Sharing Information to Improve Outcomes' An Australian Government funded initiative Review of Recovery Measures.*

Cavelti, M., Kvrgic, S., Beck, E. M., Kossowsky, J., & Vauth, R. (2012). Assessing recovery from schizophrenia as an individual process. A review of self-report

instruments. *European Psychiatry*, *27*(1), 19-32. http://doi.org/10.1016/j.eurpsy.2011.01.007

Cechnicki, A., Angermeyer, M. C., & Bielan´ska, A. (2011). Anticipated and experienced stigma among people with schizophrenia: Its nature and correlates. *Social Psychiatry and Psychiatric Epidemiology*, *46*, 643– 650. http://dx.doi.org/10.1007/s00127-010-0230-2

Chambers, C. (2017). *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice.* Princeton University Press.

Chamberlin, J. (1990). The ex-patients' movement: Where we've been and where we're going. *The Journal of Mind and Behavior*, *11*(3/4), 323–336. http://www.jstor.org/stable/43854095

Chang, S., Ong, H. L., Seow, E., Chua, B. Y., Abdin, E., Samari, E., ... & Subramaniam, M. (2017). Stigma towards mental illness among medical and nursing students in Singapore: A cross-sectional study. *BMJ Open*, *7*(12), e018099. http://dx.doi.org/10.1136/bmjopen-2017-018099

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Chester, P., Ehrlich, C., Warburton, L., Baker, D., Kendall, E., & Crompton, D. (2016). "What is the work of Recovery Oriented Practice? A systematic literature review". *International Journal of Mental Health Nursing*, *25*(4), 270–285. https://doi.org/10.1111/inm.12241

Chiba, R., Umeda, M., Goto, K., Miyamoto, Y., Yamaguchi, S., & Kawakami, N. (2016). Psychometric properties of the Japanese version of the Recovery Attitudes Questionnaire (RAQ) among mental health providers: A questionnaire survey. *BMC Psychiatry*, *16*(1), 1-9. https://doi.org/10.1186/s12888-016-0740-x

Christison, G. W., Haviland, M. G., & Riggs, M. L. (2002). The medical condition regard scale: measuring reactions to diagnoses. *Academic medicine : journal of the Association of American Medical Colleges*, *77*(3), 257–262. https://doi.org/10.1097/00001888-200203000-00017

Clarke, C., Lumbard, D., Sambrook, S., & Kerr, K. (2016). What does recovery mean to a forensic mental health patient? A systematic review and narrative synthesis of the qualitative literature. *Journal of Forensic Psychiatry & Psychology*, *27,* 38–54. http://dx.doi.org/10.1080/1478 9949.2015.1102311

Clement, S., Jarrett, M., Henderson, C., & Thornicroft, G. (2010). Messages to use in population-level campaigns to reduce mental health-related stigma: Consensus development study. *Epidemiologia E Psichiatria Sociale*, *19*(1), 72-79. https://doi.org/10.1017/S1121189X00001627

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cohen, J. L., Abraham, K. M., Burk, J. P., & Stein, C. H. (2012). Emerging opportunities for psychologists: Joining consumers in the recovery-oriented care movement. *Professional Psychology: Research and Practice*, *43*(1), 24.

Cohen, R. J., & Swerdlik, M. (2009). *Psychological Testing and Assessment* (7th ed.). New York: McGraw-Hill Humanities/Social Sciences/Languages.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, *55*(4), 584–594. https://doi.org/10.1037/0022-006X.55.4.584

Commonwealth of Australia. (2009). *Fourth National Mental Health Plan. An Agenda for Collaborative Government Action in Mental Health 2009–2014.* Department of Health Australia. Retrieved April 19, 2022, from http://www.health.gov.au/internet/main/publishing.nsf/Content/9A5A0E8BDFC55D3BCA257BF0001C1B1C/$File/plan09v2.pdf.

Cook, J. E., Purdie-Vaughns, V., Meyer, I. H., & Busch, J. (2014). Intervening within and across levels: a multilevel approach to stigma and public health. *Social Science & Medicine (1982)*, *103*, 101–109. https://doi.org/10.1016/j.socscimed.2013.09.023

Corrigan, P. W. (2004). How stigma interferes with mental health care. *American Psychologist, 59*, 614–625. http://dx.doi.org/10.1037/0003-066X .59.7.614

Corrigan, P. W. (2005). *On the Stigma of Mental Illness: Implications for Research and Social Change*. Washington, DC: American Psychological Association Press. http://dx.doi.org/10.1037/10887-000

Corrigan, P. W., Kerr, A., & Knudsen, L. (2005). The stigma of mental illness: Explanatory models and methods for change. *Applied & Preventive Psychology, 11,* 179–190. http://dx.doi.org/10.1016/j.appsy .2005.07.001

Corrigan, P. W., Larson, J. E., & Rüsch, N. (2009). Self-stigma and the "why try" effect: Impact on life goals and evidence-based practices. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 8*(2), 75–81. https://doi.org/10.1002/j.2051-5545.2009.tb00218.x

Corrigan, P., Markowitz, F. E., Watson, A., Rowan, D., & Kubiak, M. A. (2003). An attribution model of public discrimination towards persons with mental illness. *Journal of Health and Social Behavior*, *44*(2), 162–179. https://doi.org/10.2307/1519806

Corrigan, P. W., Michaels, P. J., Vega, E., Gause, M., Watson, A. C., & Rüsch, N. (2012). Self-Stigma of Mental Illness Scale—Short form: Reliability and validity. *Psychiatry Research, 199,* 65–69. http://dx.doi.org/10.1016/j.psychres.2012.04.009

Corrigan, P. W., Morris, S., Larson, J., Rafacz, J., Wassel, A., Michaels, P., ... & Rüsch, N. (2010). Self-stigma and coming out about one's mental illness. *Journal of Community Psychology*, *38*(3), 259-275. https://doi.org/10.1002/jcop.20363

Corrigan, P. W., Morris, S. B., Michaels, P. J., Rafacz, J. D., & Rüsch, N. (2012). Challenging the public stigma of mental illness: A meta-analysis of outcome studies. *Psychiatric Services 63*(10), 963-973. http://dx.doi.org/10.1176/appi. ps.201100529

Corrigan, P. W., & Rao, D. (2012). On the self-stigma of mental illness: stages, disclosure, and strategies for change. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *57*(8), 464–469. https://doi.org/10.1177/070674371205700804

Corrigan, P. W., Rowan, D., Green, A., Lundin, R., River, P., Uphoff-Wasowski, K., ... & Kubiak, M. A. (2002). Challenging two mental illness stigmas: Personal responsibility and dangerousness. *Schizophrenia Bulletin*, *28*(2), 293-309. https://doi.org/10.1093/oxfordjournals.schbul.a006939

Corrigan, P. W., & Rüsch, N. (2002). Mental illness stereotypes and clinical care: Do people avoid treatment because of stigma? *Psychiatric Rehabilitation Skills, 6*(3), 312–334. https://doi.org/10.1080/10973430208408441

Corrigan, P. W., & Watson, A. C. (2002). The paradox of self-stigma and mental illness. *Clinical Psychology: Science and Practice, 9*, 35–53. http://dx.doi.org/10.1093/clipsy.9.1.35

Corrigan, P. W., Watson, A. C., & Barr, L. (2006). The self-stigma of mental illness: Implications for self-esteem and self-efficacy. *Journal of Social and Clinical Psychology, 25*, 875– 884. http://dx.doi.org/10.1521/jscp.2006.25.8.875

Corrigan, P. W., Watson, A. C., Warpinski, A. C., & Gracia, G. (2004). Stigmatising attitudes about mental illness and allocation of resources to mental health services. *Community Mental Health Journal, 40*, 297–307. http://dx.doi.org/10.1023/B:COMH.0000035226.19939.76

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608–630. https://doi.org/10.1037/0033-295X.96.4.608

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Crossley, N. (2006). *Contesting Psychiatry: Social Movements in Mental Health*. Routledge.

Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, *11*(3), 242-247. https://doi.org/10.1080/17437199.2015.1124240

Dalziel, L. (2001). The New Zealand disability strategy: Making a world of difference: Whakanui oranga. *Wellington: Ministry of Health*.

Davidson, L. (2012). Considering recovery as a process: Or, life is not an outcome. In A. Rudnick (Ed.), *Recovery of People with Mental Illness: Philosophical and Related Perspectives* (pp. 252–263). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/med/9780199691319 .003.0016

Davidson, L., & Roe, D. (2007). Recovery from versus recovery in serious mental illness: One strategy for lessening confusion plaguing recovery. *Journal of Mental Health, 16*, 459 - 470. https://doi.org/10.1080/15487760500339360

Davidson, L., Tondora, J., O'Connell, M. J., Kirk, T., Jr., Rockholz, P., & Evans, A. C. (2007). Creating a recovery-oriented system of behavioral health care: Moving from concept to

reality. *Psychiatric Rehabilitation Journal, 31*(1), 23–31. https://doi.org/10.2975/31.1.2007.23.31

Deegan, P. E. (1988). Recovery: The lived experience of rehabilitation. *Psychosocial Rehabilitation Journal, 11*, 11–19. http://dx.doi.org/10.1037/h0099565

Department of Health. (2001). *The Journey to Recovery: The Government's Vision for Mental Health Care*. Retrieved April 19, 2022, from http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4058900.pdf.

Destrebecq, A., Ferrara, P., Frattini, L., Pittella, F., Rossano, G., Striano, G., ... & Gambini, O. (2018). The Italian version of the Opening Minds Stigma Scale for Healthcare Providers: Validation and study on a sample of bachelor students. *Community Mental Health Journal*, *54*(1), 66-72. https://doi.org/10.1007/s10597-017-0149-0

De Ruysscher, C., Vandevelde, S., Vanderplasschen, W., De Maeyer, J., & Vanheule, S. (2017). The concept of recovery as experienced by persons with dual diagnosis: A systematic review of qualitative research from a first-person perspective. *Journal of Dual Diagnosis, 13*, 264 –279. http://dx.doi.org/10.1080/15504263.2017.1349977

de Vos, J. A., LaMarre, A., Radstaak, M., Bijkerk, C. A., Bohlmeijer, E. T., & Westerhof, G. J. (2017). Identifying fundamental criteria for eating disorder recovery: A systematic review and qualitative meta-analysis. *Journal of Eating Disorders, 5,* 34. http://dx.doi.org/10.1186/s40337- 017-0164-0

DiStefano, C., & Hess, B. (2005). Using Confirmatory Factor Analysis for Construct Validation: An Empirical Review. *Journal of Psychoeducational Assessment, 23*(3), 225–241. https://doi.org/10.1177/073428290502300303

Dueber, D. M. (2017). *Bifactor Indices Calculator: A Microsoft Excel-based Tool to Calculate Various Indices Relevant to Bifactor CFA models*. https://doi.org/10.13023/edp.tool.01

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Dyess, S. M. (2011). Faith: A concept analysis. *Journal of Advanced Nursing*, *67*(12), 2723-2731. https://doi.org/10.1111/j.1365-2648.2011.05734.x

Ellison, M. L., Belanger, L. K., Niles, B. L., Evans, L. C., & Bauer, M. S. (2018). Explication and definition of mental health recovery: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research, 45,* 91–102. http://dx.doi.org/10.1007/s10488-016- 0767-9

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1-20. https://doi.org/10.3758/s13428-021-01694-3

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*, 186–192. https://doi.org/10.1177/002224378802500207

Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 62-72. https://doi.org/10.1080/10705519609540030

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. New York, NY: Simon & Schuster Inc.

Gordon, S., Ellis, P., Gallagher, P., & Purdie, G. (2014). Service users teaching the recovery paradigm to final year medical students. A New Zealand approach. *Health Issues*, (113), 15–17. https://search.informit.org/doi/10.3316/informit.834822827421339

Gordon, S. E., Kininmonth, L. A., Newton-Howes, G., Purdie, G. L., & Gardiner, T. (2021). A comparative analysis of two alternative programmes to counter stigma and discrimination associated with mental illness delivered to medical students. *Australasian Psychiatry*, *29*(3), 357–360. https://doi.org/10.1177/1039856220971935

Feldman, D. B., & Crandall, C. S. (2007). Dimensions of mental illness stigma: What about mental illness causes social rejection? *Journal of Social and Clinical Psychology*, *26*(2), 137-154. https://doi.org/10.1521/jscp.2007.26.2.137

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, 456–465. https://doi.org/10.1177/2515245920952393

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological & Personality Science*, *8*, 370–378. https://doi.org/10.1177/1948550617693063

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble: Overfitting in the assessment of internal structure and some editorial thoughts on it [Editorial]. *European Journal of Psychological Assessment*, *33*(6), 399–402. https://doi.org/10.1027/1015-5759/a000460

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association: JMLA*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Foster, K., Withers, E., Blanco, T., Lupson, C., Steele, M., Giandinoto, J. A., & Furness, T. (2019). Undergraduate nursing students' stigma and recovery attitudes during mental health clinical placement: A pre/post-test survey study. *International Journal of Mental Health Nursing*, *28*(5), 1068-1080. https://doi.org/10.1111/inm.12634

Fox, A. B., Earnshaw, V. A., Taverna, E. C., & Vogt, D. (2018). Conceptualising and measuring mental illness stigma: The mental illness stigma framework and critical review of measures. *Stigma and Health*, *3*(4), 348–376. https://doi.org/10.1037/sah0000104

Griffiths, K. M., Carron-Arthur, B., Parsons, A., & Reid, R. (2014). Effectiveness of programs for reducing the stigma associated with mental disorders. A meta-analysis of randomized controlled trials. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *13*(2), 161–175. https://doi.org/10.1002/wps.20129

Gwarjanski, A. R., & Parrott, S. (2018). Schizophrenia in the News: The Role of News Frames in Shaping Online Reader Dialogue about Mental Illness. *Health Communication*, *33*(8), 954–961. https://doi.org/10.1080/10410236.2017.1323320

Gyamfi, N., Bhullar, N., Islam, M. S., & Usher, K. (2021). A systematic review of measures assessing mental health professionals' perspectives of recovery. *International Journal of Mental Health Nursing*, *30*(4), 847–874. https://doi.org/10.1111/inm.12895

Happell, B., Platania-Phung, C., Scholz, B., Bocking, J., Horgan, A., Manning, F., Doody, R., Hals, E., Granerud, A., Lahti, M., Pullo, J., Vatula, A., Koski, J., van der Vaart, K. J., Allon, J., Griffin, M., Russell, S., MacGabhann, L., Bjornsson, E., & Biering, P. (2019). Changing attitudes: The impact of Expert by Experience involvement in Mental Health Nursing Education: An international survey study. *International Journal of Mental Health Nursing*, *28*(2), 480–491. https://doi.org/10.1111/inm.12551

Hammer, J. H., & Toland, M. D. (2016). *Bifactor Analysis in Mplus*. Lexington, KY: University of Kentucky. Retrieved from http://sites.education.uky.edu/apslab/upcoming-events/

Hardiman, E. R. & Hodges, J. Q. (2008). Professional differences in attitudes toward and utilization of psychiatric recovery. *Families in Society*, *89*(2), 220–227. https://doi-org.ezproxy.massey.ac.nz/10.1606/1044-3894.3737

Harding, C. M., Brooks, G. W., Ashikaga, T., Strauss, J. S., & Breier, A. (1987). The Vermont longitudinal study of persons with severe mental illness, II: Long-term outcome of subjects who retrospectively met DSM-III criteria for schizophrenia. *The American Journal of Psychiatry*, *144*(6), 727–735. https://doi.org/10.1176/ajp.144.6.727

Hasson-Ohayon, I., Kravetz, S., & Lysaker, P. H. (2017). The special challenges of psychotherapy with persons with psychosis: Intersubjective metacognitive model of

agreement and shared meaning. *Clinical Psychology & Psychotherapy, 24*, 428–440. http://dx.doi.org/10.1002/cpp.2012

Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health, 103*, 813–821. http://dx.doi.org/10.2105/AJPH.2012 .301069

Hebl, M. R., Tickle, J., & Heatherton, T. F. (2000). Awkward moments in interactions between nonstigmatised and stigmatised individuals. In T. F. Heatherton, R. E. Kleck, M. R. Hbel, & J. G. Hull (Eds.), *The Social Psychology of Stigma* (pp. 275–306). New York, NY: Guilford.

Honaker, J., King, G., & Blackwell, M. (2011). "Amelia II: A Program for Missing Data." *Journal of Statistical Software*, *45*(7), 1–47. https://www.jstatsoft.org/v45/i07/

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/10705 519909540118

Hungerford, C., Dowling, M., & Doyle, K. (2015). Recovery outcome measures: Is there a place for culture, attitudes, and faith? *Perspectives in Psychiatric Care*, *51*(3), 171-179. https://doi.org/10.1111/ppc.12078

Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, *10*(1), 128-141. https://doi.org/10.1207/S15328007SEM1001_6

Jacob K. S. (2015). Recovery model of mental illness: A complementary approach to psychiatric care. *Indian Journal of Psychological Medicine*, *37*(2), 117–119. https://doi.org/10.4103/0253-7176.155605

Jacobson, N. (2001). Experiencing recovery: A dimensional analysis of recovery narratives. *Psychiatric Rehabilitation Journal*, *24*(3), 248–256. https://doi.org/10.1037/h0095087

Jaeger, M., Konrad, A., Rueegg, S., & Rabenschlag, F. (2013). Measuring recovery: Validity of the "recovery process inventory" and the "recovery attitudes questionnaire". *Psychiatry Research*, *210*(1), 363-367. https://doi.org/10.1016/j.psychres.2013.06.002

Jose, D., Ramachandra, , Lalitha, K., Gandhi, S., Desai, G., & Nagarajaiah. (2015). Consumer perspectives on the concept of recovery in schizophrenia: A systematic review. *Asian Journal of Psychiatry*, *14*, 13–18. http://dx.doi.org/10.1016/j.ajp.2015.01.006

Jones, E. E. (1984). *Social Stigma: The Psychology of Marked Relationships*. New York: W.H. Freeman.

Kane J. M. (2013). Improving patient outcomes in schizophrenia: Achieving remission, preventing relapse, and measuring success. *The Journal of Clinical Psychiatry*, *74*(9), e18. https://doi.org/10.4088/JCP.12117tx1cc

Kassam, A., Papish, A., Modgill, G., & Patten, S. (2012). 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, 12*(1), 1-12. https://doi.org/10.1186/1471-244X-12-62

Kelley, K. (2017). MBESS (Version 4.8.1) [computer software and manual]. Accessible from http://cran.r-project.org.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling, 12*(3), 368-390. [Sample size formula, Non-centrality parameter calculation]

Klin, A., & Lemish, D. (2008). Mental disorders stigma in the media: Review of studies on production, content, and influences. *Journal of Health Communication*, *13*(5), 434–449. https://doi.org/10.1080/10810730802198813

Knaak, S., Modgill, G., & Patten, S. B. (2014). Key Ingredients of Anti-Stigma Programs for Health Care Providers: A Data Synthesis of Evaluative Studies. *The Canadian Journal of Psychiatry, 59*(1_suppl), 19–26. https://doi.org/10.1177/070674371405901S06

Knekta, E., Runyon, C., & Eddy, S. (2019). One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. *CBE Life Sciences Education*, *18*(1), rm1. https://doi.org/10.1187/cbe.18-04-0064

Koenig H. G. (2004). Religion, spirituality, and medicine: Research findings and implications for clinical practice. *Southern Medical Journal*, *97*(12), 1194–1200. https://doi.org/10.1097/01.SMJ.0000146489.21837.CE

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? *Sociological Methods & Research, 41*(1), 124–167. https://doi.org/10.1177/0049124112442138

Korlén, S., Richter, A., Amer-Wåhlin, I., Lindgren, P., & von Thiele Schwarz, U. (2018). The development and validation of a scale to explore staff experience of Governance of Economic Efficiency and Quality (GOV-EQ) of Health Care. *BMC Health Services Research*, *18*(1), 963. https://doi.org/10.1186/s12913-018-3765-7

Kruger, A. (2000). Schizophrenia: Recovery and hope. *Psychiatric Rehabilitation Journal*, *24*(1), 29–37. https://doi.org/10.1037/h0095126

Kruyen, P. M., Emons, W. H., & Sijtsma, K. (2012). Test length and decision quality in personnel selection: When is short too short? *International Journal of Testing*, *12*(4), 321-344. https://doi.org/10.1080/15305058.2011.643517

Lakeman, R. (2013). Talking science and wishing for miracles: Understanding cultures of mental health practice. *International Journal of Mental Health Nursing*, *22*(2), 106–115. https://doi.org/10.1111/j.1447-0349.2012.00847.x

Law, H., & Morrison, A. P. (2014). Recovery in psychosis: A Delphi study with experts by experience. *Schizophrenia Bulletin*, *40*(6), 1347–1355. https://doi.org/10.1093/schbul/sbu047

Law, H., Morrison, A., Byrne, R., & Hodson, E. (2012). Recovery from psychosis: A user informed review of self-report instruments for measuring recovery. *Journal of Mental Health, 21,* 192–207. http://dx.doi.org/10.3109/09638237.2012.670885

Leamy, M., Bird, V., Le Boutillier, C., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: Systematic review and narrative synthesis. *The British Journal of Psychiatry, 199,* 445–452. http://dx.doi.org/10.1192/bjp.bp.110.083733

Leete, E. (1989). How I perceive and manage my illness. *Schizophrenia Bulletin, 15,* 197-200.

Leonhardt, B. L., Huling, K., Hamm, J. A., Roe, D., Hasson-Ohayon, I., McLeod, H. J., & Lysaker, P. H. (2017). Recovery and serious mental illness: A review of current clinical and research paradigms and future directions. *Expert Review of Neurotherapeutics*, *17*(11), 1117–1130. https://doi.org/10.1080/14737175.2017.1378099

Le Boutillier, C., Slade, M., Lawrence, V., Bird, V. J., Chandler, R., Farkas, M., Harding, C., Larsen, J., Oades, L. G., Roberts, G., Shepherd, G., Thornicroft, G., Williams, J., & Leamy, M. (2015). Competing priorities: Staff perspectives on supporting recovery. *Administration and Policy in Mental Health*, *42*(4), 429–438. https://doi.org/10.1007/s10488-014-0585-x

Liberman, R. P., & Kopelowicz, A. (2005). Recovery from schizophrenia: A concept in search of research. *Psychiatric Services, 56,* 735–742. http://dx.doi.org/10.1176/appi.ps.56.6.735 McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Erlbaum.

Lin, J. (2021). *Introduction to Structural Equation Modeling (SEM) in R with lavaan.* OARC Stats. Retrieved 2021, from https://stats.oarc.ucla.edu/r/seminars/rsem/

Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. *American Sociological Review, 52*(1), 96–112. https://doi.org/10.2307/2095395

Link, B. G., & Phelan, J. C. (2001). Conceptualising stigma. *Annual Review of Sociology, 27,* 363–385. http://doi.org/10.1146/annurev .soc.27.1.363

Link, B. G., Yang, L. H., Phelan, J. C., & Collins, P. Y. (2004). Measuring mental illness stigma. *Schizophrenia Bulletin*, *30*(3), 511–541. https://doi.org/10.1093/oxfordjournals.schbul.a007098

Livingston, J. D., & Boyd, J. E. (2010). Correlates and consequences of internalised stigma for people living with mental illness: A systematic review and meta-analysis. *Social Science & Medicine, 71*, 2150–2161. http://dx.doi.org/10.1016/j.socscimed.2010.09.030

Livingston, J. D., Cianfrone, M., Korf-Uzan, K., & Coniglio, C. (2014). Another time point, a different story: one year effects of a social media intervention on the attitudes of young people towards mental health issues. *Social Psychiatry and Psychiatric Epidemiology*, *49*(6), 985–990. https://doi.org/10.1007/s00127-013-0815-7

Luigi, M., Rapisarda, F., Corbière, M., De Benedictis, L., Bouchard, A. M., Felx, A., Miglioretti, M., Abdel-Baki, A., & Lesage, A. (2020). Determinants of mental health professionals'

attitudes towards recovery: A review. *Canadian Medical Education Journal*, *11*(5), e62–e73. https://doi.org/10.36834/cmej.61273

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637. https://doi.org/10.1207/S15327906MBR3604_06

Machin, D., Campbell, M. J., Tan, S. B., & Tan, S. H. (2009). Comparing two independent groups for binary data. *Sample Size Tables for Clinical Studies*. edn. Oxford: Wiley-Blackwell, 30-41. https://doi.org/10.1002/9781444300710

Mancini, M. A., Hardiman, E. R., & Lawson, H. A. (2005). Making Sense of It All: Consumer Providers' Theories about Factors Facilitating and Impeding Recovery from Psychiatric Disabilities. *Psychiatric Rehabilitation Journal*, *29*(1), 48–55. https://doi.org/10.2975/29.2005.48.55

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519-530.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*, 537-563. https://doi.org/10.1146/annurev.psych.59.103006.093735

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review: An Official Journal of The Society for Personality and Social Psychology, Inc*, *15*(1), 28–50. https://doi.org/10.1177/1088868310366253

McDonald, R.P. (1999). *Test Theory: A Unified Treatment*. Hillsdale: Erlbaum

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Mental Health Commission of Canada. (2015). *Updated International Review of Recovery Guidelines. Final report*. Retrieved April 19, 2022, from http://www.mentalhealthcommission.ca/English/system/files/private/t/FINAL%20Updated%20MHCC%20Recovery%20Guidelines%202014%2001%2015.pdf.

Michaels, P. J., López, M., Rüsch, N., & Corrigan, P. W. (2017). Constructs and concepts comprising the stigma of mental illness. *Psychology, Society & Education*, *4*(2), 183-194. https://doi.org/10.25115/psye.v4i2.490

Ministry of Health. (2012). *Rising to the Challenge: The Mental Health and Addiction Service Development Plan 2012–2017*. Ministry of Health NZ. Retrieved from http://www.health.govt.nz/publication/rising-challenge-mental-health-and-addiction-service-development-plan-2012-2017

Ministry of Health. (2016). *Annual Report for the Year Ended 30 June 2016: Ministry of Health.* Wellington: Ministry of Health. Retrieved from https://www.health.govt.nz/system/files/documents/publications/annual-report-year-ended-30-june-2016-oct16.pdf

Ministry of Health. (2019). *Annual Update of Key Results 2018/19: New Zealand Health Survey*. Retrieved from https://www.health.govt.nz/publication/annual-update-key-results-2018-19-new-zealand-health-survey

Modgill, G., Patten, S. B., Knaak, S., Kassam, A., & Szeto, A. C. (2014). Opening minds Stigma Scale for Health Care Providers (OMS-HC): Examination of psychometric properties and responsiveness. *BMC Psychiatry*, *14*(1), 1-10. https://doi.org/10.1186/1471-244X-14-120

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Morant N. (2006). Social representations and professional knowledge: The representation of mental illness among mental health practitioners. *The British Journal of Social Psychology*, *45*(Pt 4), 817–838. https://doi.org/10.1348/014466605x810376

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116-139. https://doi.org/10.1080/10705511.2014.961800

Morrison, A. P., Law, H., Barrowclough, C., Bentall, R. P., Haddock, G., Jones, S. H., Kilbride, M., Pitt, E., Shryane, N., Tarrier, N., Welford, M., & Dunn, G. (2016). *Psychological*

Approaches to Understanding and Promoting Recovery in Psychosis and Bipolar Disorder: A Mixed-Methods Approach. NIHR Journals Library.

Mötteli, S., Horisberger, R., Lamster, F., Vetter, S., Seifritz, E., & Jäger, M. (2019). More optimistic recovery attitudes are associated with less stigmatization of people with mental illness among healthcare professionals working on acute and semi-acute psychiatric wards. *Psychiatric Quarterly*, *90*(3), 481-489. https://doi.org/10.1007/s11126-019-09642-3

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, Article 0021. https://doi.org/10.1038/s41562-016-0021

Nelson, T. D. (Ed.), (2009). *Handbook of Prejudice, Stereotyping, and Discrimination*. New York, NY: Psychology Press.

Newlin, K., Dyess, S., Allard, E., Chase, S., & D'Eramo Melkus, G. (2012). A methodological review of faith-based health promotions literature: Advancing the science to expand delivery of diabetes education to Black Americans. *Journal of Religion & Health*, *51*(4), 1075–1097. https://doi.org/10.1007/s10943-011-9481-9

Newton-Howes, G., Senior, J., Beaglehole, B., Purdie, G. L., & Gordon, S. E. (2021). Does a comprehensive service user-led education programme effect more positive attitudes towards recovery and less stigmatising attitudes towards people with lived experience of mental distress in medical students? A comparative cohort study. *The Australian and*

New Zealand Journal of Psychiatry, *55*(9), 903–910. https://doi.org/10.1177/0004867420987886

New Zealand Government. (2018). *He Ara Oranga: Report of the Government Inquiry into Mental Health and Addiction.* Wellington, New Zealand: New Zealand Government. Retrieved from https://mentalhealth.inquiry.govt.nz/inquiry-report/he-ara-oranga/

New Zealand Mental Health Commission. (1998). *Blueprint for Mental Health Services in New Zealand*. Retrieved from https://www.moh.govt.nz/notebook/nbbooks.nsf/0/0E6493ACAC236A394C25678D000BEC3C/%24file/Blueprint_for_mental_health_services.pdf

New Zealand Mental Health Commission (2001). *Recovery Competencies for New Zealand Mental Health Workers.* www.mhc.govt.nz/publications/2001/Recovery_Competencies.pdf

Ng, F. Y. Y., Bourke, M. E., & Grenyer, B. F. (2016). Recovery from borderline personality disorder: A systematic review of the perspectives of consumers, clinicians, family, and carers. *PLoS ONE, 11,* e0160515. http://dx.doi.org/10.1371/journal.pone.0160515

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*, 615–631. https://doi.org/10.1177/1745691612459058

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

O'Hagan, M. (2009). Leadership for empowerment and equality: A proposed model for mental health user/survivor leadership. *International Journal of Leadership in Public Services*. https://doi.org/10.5042/ijlps.2010.0110

O'Hagan, M., Reynolds, P., & Smith, C. (2012). Recovery in New Zealand: An evolving concept? *International Review of Psychiatry*, *24*(1), 56-63.

Orben, A., Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*, 173–182. https://doi.org/10.1038/s41562-018-0506-1

Őri, D., Rózsa, S., Szocsics, P., Simon, L., Purebl, G., & Győrffy, Z. (2020). Factor structure of the Opening Minds Stigma Scale for Health Care Providers and psychometric properties of its Hungarian version. *BMC Psychiatry*, *20*(1), 1-9. https://doi.org/10.1186/s12888-020-02902-8

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. (Version 4.1.0). Retrieved from https://www.R-project.org/

Piat, M., & Lal, S. (2012). Service providers' experiences and perspectives on recovery-oriented mental health system reform. *Psychiatric Rehabilitation Journal*, *35*(4), 289–296. https://doi.org/10.2975/35.4.2012.289.296

Piat, M., & Sabetti, J. (2009). The development of a recovery-oriented mental health system in Canada: What the experience of Commonwealth countries tells us. *Canadian Journal of Community Mental Health*, *28*(2), 17-33. https://doi/10.7870/cjcmh-2009-0020

Piat, M., Sabetti, J., Couture, A., Sylvestre, J., Provencher, H., Botschner, J., & Stayner, D. (2009). What does recovery mean for me? Perspectives of Canadian mental health consumers. *Psychiatric Rehabilitation Journal, 32*(3), 199–207. https://doi.org/10.2975/32.3.2009.199.207

Pietrus, M. (2013). *Opening Minds Interim Report*. Mental Health Commission of Canada, Calgary. Retrieved from http://www.mentalhealthcommission.ca/English/initiatives-and-projects/opening-minds/opening-minds-interim-report.

Pilgrim D. (2008). 'Recovery' and current mental health policy. *Chronic Illness*, *4*(4), 295–304. https://doi.org/10.1177/1742395308097863

Pincus, H. A., Spaeth-Rublee, B., Sara, G., Goldner, E. M., Prince, P. N., Ramanuj, P., Gaebel, W., Zielasek, J., Großimlinghaus, I., Wrigley, M., van Weeghel, J., Smith, M., Ruud, T., Mitchell, J. R., & Patton, L. (2016). A review of mental health recovery programs in selected industrialized countries. *International Journal of Mental Health Systems*, *10*, 73. https://doi.org/10.1186/s13033-016-0104-4

Polit D. F. (2014). Getting serious about test-retest reliability: A critique of retest research and some recommendations. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *23*(6), 1713–1720. https://doi.org/10.1007/s11136-014-0632-9

Preckel F., Brunner M. (2017) Nomological Nets. In: Zeigler-Hill V., Shackelford T. (Eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_1334-1

Prolific. (2022a). *What are the Advantages and Limitations of an Online Sample?* Prolific. Retrieved 2022, from https://researcher-help.prolific.co/hc/en-gb/articles/360009501473-What-are-the-advantages-and-limitations-of-an-online-sample-

Prolific. (2022b). *How do I Prescreen for Participants who are Fluent in Multiple Languages?* Prolific. Retrieved 2022, from https://researcher-help.prolific.co/hc/en-gb/articles/360015786399-How-do-I-prescreen-for-participants-who-are-fluent-in-multiple-languages-#

Prolific. (2022c). *Prolific · Quickly Find Research Participants You Can Trust.* Prolific. Retrieved 2022, from https://www.prolific.co/

Pryor, J. B., & Reeder, G. D. (2011). HIV-related stigma. In J. C. Hall, B. J. Hall & C. J. Cockerell (Eds.), *HIV/AIDS in the Post-HAART Era: Manifestations, Treatment, and Epidemiology* (pp. 790–806). Shelton, CT: PMPH-USA.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review: DR*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Quinn, D. M., & Earnshaw, V. A. (2011). Understanding concealable stigmatised identities: The role of identity in psychological, physical, and behavioral outcomes. *Social Issues and Policy Review*, *5*, 160–190. http://dx.doi.org/10.1111/j.1751-2409.2011.01029.x

Rabenschlag, F., Konrad, A., Rueegg, S., & Jaeger, M. (2014). A recovery-oriented approach for an acute psychiatric ward: Is it feasible and how does it affect staff

satisfaction? *Psychiatric Quarterly*, *85*(2), 225-239. https://doi.org/10.1007/s11126-013-9285-z

Rakov, T., & Marcoulides, G. A. (2000). *A First Course in Structural Equation Modeling*. Mahwah, NJ: Erlbaum.

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Differences, 35*(4), 212–220. https://doi.org/10.1027/1614-0001/a000141

Rao, D., Kekwaletswe, T. C., Hosek, S., Martinez, J., & Rodriguez, F. (2007). Stigma and social barriers to medication adherence with urban youth living with HIV. *AIDS care*, 19(1), 28-33. https://doi.org/10.1080/09540120600652303

Rao, D., Elshafei, A., Nguyen, M., Hatzenbuehler, M. L., Frey, S., & Go, V. F. (2019). A systematic review of multi-level stigma interventions: State of the science and future directions. *BMC Medicine, 17*(1), 1-11. https://doi.org/10.1186/s12916-018-1244-y

Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment, 84*(3), 228–238. https://doi.org/10.1207/s15327752jpa8403_02

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *16 Suppl 1*, 19–31. https://doi.org/10.1007/s11136-007-9183-7

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective. Educational and Psychological Measurement, *73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Revelle, W. (2021). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.9, https://CRAN.R-project.org/package=psych

Revelle, W., & Condon, D. (2018). Reliability from α to ω: A tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/2y3w9

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145–154. https://doi.org/10.1007/s11336-008-9102-z

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*(1), 37–47. https://doi.org/10.1007/BF02294204

Rissmiller, D. J., & Rissmiller, J. H. (2006). Open forum: Evolution of the antipsychiatry movement into mental health consumerism. *Psychiatric Services, 57*(6), 863-866.

Ritsher, J. B., Otilingam, P. G., & Grajales, M. (2003). Internalised stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Research, 121*, 31– 49. http://dx.doi.org/10.1016/j.psychres.2003.08.008

Roberts, G., & Wolfson, P. (2004). The rediscovery of recovery: Open to all. *Advances in Psychiatric Treatment, 10*(1), 37-48. https://doi.org/10.1192/apt.10.1.37

Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*(3), 523-534. https://doi.org/10.1016/j.labeco.2009.04.005

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. URL http://www.jstatsoft.org/v48/i02/

Rössler W. (2016). The stigma of mental disorders: A millennia-long history of social exclusion and prejudices. *EMBO reports*, *17*(9), 1250–1253. https://doi.org/10.15252/embr.201643041

Rüsch, N., Corrigan, P. W., Wassel, A., Michaels, P., Olschewski, M., Wilkniss, S., & Batia, K. (2009). A stress-coping model of mental illness stigma: I. Predictors of cognitive stress

appraisal. *Schizophrenia Research*, *110*(1-3), 59–64. https://doi.org/10.1016/j.schres.2009.01.006

Russell, D. W. (2002). In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*(12), 1629–1646. https://doi.org/10.1177/014616702237645

Salgado, J. D., Deane, F. P., Crowe, T. P., & Oades, L. G. (2010). Hope and improvements in mental health service providers' recovery attitudes following training. *Journal of Mental Health (Abingdon, England)*, *19*(3), 243–248. https://doi.org/10.3109/09638230903531126

Salsman, J., Garcia, S., Lai, J., & Cella, D. (2012). Have a little faith: The impact of illness on positive and negative aspects of faith. *Psycho-Oncology*, *21*(12), 1357–1361. https://doi.org/10.1002/pon.2051

Salyers, M. P., Rollins, A. L., McGuire, A. B., & Gearhart, T. (2009). Barriers and facilitators in implementing illness management and recovery for consumers with severe mental illness: Trainee perspectives. *Administration and Policy in Mental Health*, *36*(2), 102–111. https://doi.org/10.1007/s10488-008-0200-0

Sapag, J. C., Klabunde, R., Villarroel, L., Velasco, P. R., Álvarez, C., Parra, C., ... & Corrigan, P. (2019). Validation of the Opening Minds Scale and patterns of stigma in Chilean primary health care. *PloS One*, *14*(9), e0221825. https://doi.org/10.1371/journal.pone.0221825

Sassenberg, K., & Ditrich, L. (2019). Research in Social Psychology Changed Between 2011 and 2016: Larger Sample Sizes, More Self-Report Measures, and More Online Studies. *Advances in Methods and Practices in Psychological Science*, 107–114. https://doi.org/10.1177/2515245919838781

Scheyett, A., DeLuca, J., & Morgan, C. (2013). Recovery in Severe Mental Illnesses: A Literature Review of Recovery Measures. *Social Work Research*, *37*(3), 286–303. http://www.jstor.org/stable/42659870

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350–353. https://doi.org/10.1037/1040- 3590.8.4.350

Schumacker, R. E., & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling.* Lawrence Erlbaum Associates, Inc.

Schulze B. (2007). Stigma and mental health professionals: A review of the evidence on an intricate relationship. *InternationalRreview of Psychiatry (Abingdon, England)*, *19*(2), 137–155. https://doi.org/10.1080/09540260701278929

Schwartz, C. E., & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes*, *2*(1), 1-11. https://doi.org/10.1186/1477-7525-2-16

Schwartz, C. E., & Sprangers, M. A. G. (2000). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. In C. E. Schwartz & M. A. G. Sprangers (Eds.), *Adaptation to Changing Health: Response Shift in Quality-of-*

life Research (pp. 81–107). American Psychological Association. https://doi.org/10.1037/10382-006

Seikkula, J., Aaltonen, J., Alakare, B., Haarakangas, K., Keränen, J., & Lehtinen, K. (2006). Five-year experience of first-episode nonaffective psychosis in open-dialogue approach: Treatment principles, follow-up outcomes, and two case studies. *Psychotherapy Research*, *16*(02), 214-228. https://doi.org/10.1080/10503300500268490

Serafini, G., Pompili, M., Haghighat, R., Pucci, D., Pastina, M., Lester, D., Angeletti, G., Tatarelli, R., & Girardi, P. (2011). Stigmatization of schizophrenia as perceived by nurses, medical doctors, medical students and patients. Journal of psychiatric and mental health nursing, *18*(7), 576–585. https://doi.org/10.1111/j.1365-2850.2011.01706.x

Sevilla, C. G., Ochave, J. A., Punsalan, T. G., Regala, B. P., & Uriarte, G. G. (1993). *Pengantar metode penelitian*. Jakarta: Universitas Indonesia.

Sharac, J., McCrone, P., Clement, S., & Thornicroft, G. (2010). The economic impact of mental health stigma and discrimination: A systematic review. *Epidemiologia e Psichiatria Sociale*, *19*(3), 223–232. https://doi.org/10.1017/s1121189x00001159

Shepherd, A., Doyle, M., Sanders, C., & Shaw, J. (2016). Personal recovery within forensic settings—Systematic review and meta-synthesis of qualitative methods studies. *Criminal Behaviour and Mental Health*, *26*, 59–75. http://dx.doi.org/10.1002/cbm.1966

Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using Structural Equation Modeling to Assess Functional Connectivity in the Brain Power and Sample Size

Considerations. *Educational and Psychological Measurement*, https://doi.org/10.1177/0013164414525397

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sijtsma, K. (2016). Playing with data—Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika*, *81*, 1–15. https://doi.org/10.1007/s11336-015-9446-0

Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Sirey, J. A., Bruce, M. L., Alexopoulos, G. S., Perlick, D. A., Raue, P., Friedman, S. J., & Meyers, B. S. (2001). Perceived stigma as a predictor of treatment discontinuation in young and older outpatients with depression. *American Journal of Psychiatry, 158,* 479–481. http://dx.doi.org/10.1176/appi.ajp.158.3.479

Sklar, M., Groessl, E. J., O'Connell, M., Davidson, L., & Aarons, G. A. (2013). Instruments for measuring mental health recovery: A systematic review. *Clinical Psychology Review, 33,* 1082–1095. http://dx.doi.org/10.1016/j.cpr.2013.08.002

Slade, M. (2013). *100 Ways to Support Recovery*, 2nd edition. London: Rethink Mental Illness. Retrieved from

https://www.researchintorecovery.com/files/100%20Ways%20to%20support%20recovery%202nd%20edition.pdf

Slade, M., Leamy, M., Bacon, F., Janosik, M., Le Boutillier, C., Williams, J., & Bird, V. (2012). International differences in understanding recovery: Systematic review. *Epidemiology and Psychiatric Sciences, 21*, 353–364. http://dx.doi.org/10.1017/S2045796012000133

Slade, M., & Longden, E. (2015). Empirical evidence about recovery and mental health. *BMC Psychiatry*, *15*, 285. https://doi.org/10.1186/s12888-015-0678-4

Spaniol, L. & Koehler, M. (Eds). (1994). *The Experience of Recovery.* Boston, MA: Boston University, Center for Psychiatric Rehabilitation.

Spaniol, L., Wewiorski, N. J., Gagne, C., & Anthony, W. A. (2002). The process of recovery from schizophrenia. *International Review of Psychiatry, 14*(4), 327-336. https://doi.org/10.1080/0954026021000016978

Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Vol. 82). Sage.

Stacy, M. A., & Rosenheck, R. (2019). The association of recovery orientation and stigmatizing beliefs. *Journal of Mental Health*, *28*(3), 276-281. https://doi.org/10.1080/09638237.2017.1417573

Stevens Manser, S., Chubinsky, K., & Kuhn, W. (2018). *Recovery Outcome Measures to Advance Recovery Oriented Systems of Care*. Texas Institute for Excellence in Mental Health, School of Social Work, University of Texas at Austin.

Steyer, R., Geiser, C., & Fiege, C. (2012). Latent state-trait models. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA Handbook of Research Methods in Psychology, Vol. 3. Data Analysis and Research Publication* (pp. 291–308). American Psychological Association. https://doi.org/10.1037/13621-014

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52(4), 589–617. https://doi.org/10.1007/BF02294821

Stuart, H. (2006). Mental illness and employment discrimination. *Current Opinion in Psychiatry*, *19*(5), 522-526. https://doi.org/10.1097/01.yco.0000238482.27270.5d

Stuart, H., Chen, S. P., Christie, R., Dobson, K., Kirsh, B., Knaak, S., Koller, M., Krupa, T., Lauria-Horner, B., Luong, D., Modgill, G., Patten, S. B., Pietrus, M., Szeto, A., & Whitley, R. (2014). Opening minds in Canada: Background and rationale. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *59*(10 Suppl 1), S8–S12. https://doi.org/10.1177/070674371405901s04

Stuart, S. R., Tansey, L., & Quaile, E. (2017). What we talk about when we talk about recovery: A systematic review and best-fit framework synthesis of qualitative literature. *Journal of Mental Health*, *26,* 291–304. http://dx.doi.org/10.1080/09638237.2016.1222056

Subramaniam, M., Abdin, E., Picco, L., Pang, S., Shafie, S., Vaingankar, J., . . . Chong, S. (2017). Stigma towards people with mental disorders and its components – a perspective from multi-ethnic Singapore. *Epidemiology and Psychiatric Sciences*, *26*(4), 371-382. https://doi.org/10.1017/S2045796016000159

Substance Abuse and Mental Health Services Administration. (2021). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2020 National Survey on Drug Use and Health* (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Sutton, R. & French, P. (2019), "Challenges of measuring the influence of the Recovery Academy upon health professionals: Considerations for quantitative research", *The Journal of Mental Health Training, Education and Practice*, Vol. 14 No. 3, pp. 149-155. https://doi.org/10.1108/JMHTEP-04-2018-0027

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481-498). Boston, MA: Pearson.

Taylor, J. E., & Gordon, S. E. (2020). Evaluating service user-led teaching of mental health recovery concepts in clinical psychology training. Training and Education in Professional Psychology. *Advance Online Publication*. https://doi.org/10.1037/tep0000377

Telles-Correia, D., Saraiva, S., & Gonçalves, J. (2018). Mental disorder - The need for an accurate definition. *Frontiers in Psychiatry*, *9*, 64. https://doi.org/10.3389/fpsyt.2018.00064

Tew, J., Ramon, S., Slade, M., Bird, V., Melton, J., & Le Boutillier, C. (2012). Social factors and recovery from mental health difficulties: A review of the evidence. *The British Journal of Social Work*, *42*(3), 443-460. https://doi.org/10.1093/bjsw/bcr076

Thornicroft, G. (2006). *Shunned: Discrimination Against People with Mental Illness* (Vol. 301). Oxford: Oxford university press.

Thornicroft, G., & Kassam, A. (2008). Public attitudes, stigma and discrimination against people with mental illness. In C. Morgan, K. McKenzie, & P. Fearon (Eds.), *Society and psychosis* (pp. 179–197). Cambridge University Press. https://doi.org/10.1017/CBO9780511544064.012

Thornicroft, G., Rose, D., & Kassam, A. (2007). Discrimination in health care against people with mental illness. *International Review of Psychiatry (Abingdon, England)*, *19*(2), 113–122. https://doi.org/10.1080/09540260701278937

Tooth, B., Kalyanasundaram, V., Glover, H., & Momenzadah, S. (2003). Factors Consumers Identify as Important to Recovery from Schizophrenia. *Australasian Psychiatry*, *11*(1_suppl), S70–S77. https://doi.org/10.1046/j.1440-1665.11.s1.1.x

Topor, A., Borg, M., Di Girolamo, S., & Davidson, L. (2011). Not Just an Individual Journey: Social Aspects of Recovery. *International Journal of Social Psychiatry*, *57*(1), 90–99. https://doi.org/10.1177/0020764009345062

Turner, J. C., & TenHoor, W. J. (1978). The NIMH Community Support Program: Pilot approach to a needed social reform. *Schizophrenia Bulletin*, *4*(3), 319–348. https://doi.org/10.1093/schbul/4.3.319

Unzicker, R. (1989). On my own: A personal journey through madness 8c re-emergence. *Psychosocial Rehabilitation Journal*, *23*(1), 71-77.

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1-67. https://doi.org/0.18637/jss.v045.i03

Van Eck, R. M., Burger, T. J., Vellinga, A., Schirmbeck, F., & de Haan, L. (2018). The relationship between clinical and personal recovery in patients with schizophrenia spectrum disorders: A systematic review and meta-analysis. *Schizophrenia Bulletin*, *44*, 631– 642. http://dx.doi.org/10 .1093/schbul/sbx088

van der Maas, M., Stuart, H., Patten, S. B., Lentinello, E. K., Bobbili, S. J., Mann, R. E., Hamilton, H. A., Sapag, J. C., Corrigan, P., & Khenti, A. (2018). Examining the Application of the Opening Minds Survey in the Community Health Centre Setting. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *63*(1), 30–36. https://doi.org/10.1177/0706743717719079

van Weeghel, J., van Zelst, C., Boertien, D., & Hasson-Ohayon, I. (2019). Conceptualisations, assessments, and implications of personal recovery in mental illness: A scoping review of systematic reviews and meta-analyses. *Psychiatric Rehabilitation Journal*, *42*(2), 169–181. https://doi.org/10.1037/prj0000356

Volkow, N. D., Gordon, J. A., & Koob, G. F. (2021). Choosing appropriate language to reduce the stigma around mental illness and substance use disorders. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *46*(13), 2230–2232. https://doi.org/10.1038/s41386-021-01069-4

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R., Ahn, S. Y., Ali, M. K., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., … Memish, Z. A. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)*, *380*(9859), 2163–2196. https://doi.org/10.1016/S0140-6736(12)61729-2

Vuckovich, P. K., & Artinian, B. M. (2005). Justifying coercion. *Nursing Ethics*, *12*(4), 370-380.

Wahl, O. F. (2012). Stigma as a barrier to recovery from mental illness. *Trends in Cognitive Sciences*, *16*(1), 9-10. https://doi.org/10.1016/j.tics.2011.11.002

Wang, J., Fick, G., Adair, C., & Lai, D. (2007). Gender specific correlates of stigma toward depression in a Canadian general population sample. Journal of Affective Disorders, *103*(1-3), 91-97. https://doi.org/10.1016/j.jad.2007.01.010

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

White, W. L. (2007). *The New Recovery Advocacy Movement in America*. *Addiction*, *102*(5), 696-703. https://doi.org/10.1111/j.1360-0443.2007.01808.x

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. In K. H. Trzesniewski, M. B. Donnellan, & R. E. Lucas (Eds.), *Secondary Data Analysis* (pp. 39-61), Washington, DC: American Psychological Association.

Wieland, A., Durach, C. F., Kembro, J., & Treiblmaier, H. (2017). Statistical and judgmental criteria for scale purification. *Supply Chain Management: An International Journal*. https://doi.org/10.1108/SCM-07-2016-0230

Williams, J., Leamy, M., Bird, V., Harding, C., Larsen, J., Le Boutillier, C., Oades, L., & Slade, M. (2012). Measures of the recovery orientation of mental health services: Systematic review. *Social Psychiatry and Psychiatric Epidemiology*, *47*(11), 1827–1835. https://doi.org/10.1007/s00127-012-0484-y

Williams, D. R., Neighbors, H. W., & Jackson, J. S. (2003). Racial/ethnic discrimination and health: Findings from community studies. *American Journal of Public Health, 93,* 200–208. http://dx.doi.org/10.2105/AJPH .93.2.200

Wilrycx, G., Croon, M. A., van den Broek, A., & van Nieuwenhuizen, C. (2012). Psychometric properties of three instruments to measure recovery. *Scandinavian Journal of Caring Sciences*, *26*(3), 607-614. https://doi.org/10.1111/j.1471-6712.2011.00957.x

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety.

Educational and Psychological Measurement, *73*(6), 913-934. https://doi.org/10.1177/0013164413495237

Wood, L., & Alsawy, S. (2018). Recovery in psychosis from a service user perspective: A systematic review and thematic synthesis of current qualitative evidence. *Community Mental Health Journal*, *54*(6), 793-804. https://doi.org/10.1007/s10597-017-0185-9

Yamada Y. (2018). How to Crack Pre-registration: Toward Transparent and Open Science. *Frontiers in Psychology*, *9*, 1831. https://doi.org/10.3389/fpsyg.2018.01831

Yang, L. H., Kleinman, A., Link, B. G., Phelan, J. C., Lee, S., & Good, B. (2007). Culture and stigma: Adding moral experience to stigma theory. *Social Science & Medicine*, *64*, 1524–1535. https://doi.org/10.1016/j.socscimed.2006.11.013

Yang, L. H., Thornicroft, G., Alvarado, R., Vega, E., & Link, B. G. (2014). Recent advances in cross-cultural measurement in psychiatric epidemiology: Utilizing 'what matters most'to identify culture-specific aspects of stigma. *International Journal of Epidemiology*, *43*(2), 494-510. https://doi.org/10.1093/ije/dyu039

Yap, M. B., Mackinnon, A., Reavley, N., & Jorm, A. F. (2014). The measurement properties of stigmatizing attitudes towards mental disorders: Results from two community surveys. *International Journal of Methods in Psychiatric Research*, *23*(1), 49-61. https://doi.org/10.1002/mpr.1433

Ziegler, M. (2014). Comments on item selection procedures [Editorial]. *European Journal of Psychological Assessment*, *30*(1), 1–2. https://doi.org/10.1027/1015-5759/a000196

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes [Editorial]. *European Journal of Psychological Assessment*, *31*(4), 231–237. https://doi.org/10.1027/1015-5759/a000309

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them [Editorial]. *Journal of Individual Differences*, *35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148

Zuaboni, G., Elmer, T., Rabenschlag, F., Heumann, K., Jaeger, S., Kozel, B., Mahlke, C. I., Theodoridou, A., Jaeger, M., & Rüsch, N. (2021). Psychometric evaluation of the German version of the Opening Minds Stigma Scale for Health Care Providers (OMS-HC). *BMC psychology*, *9*(1), 86. https://doi.org/10.1186/s40359-021-00592-9

Appendix A – Survey Flow, Information Sheet and Survey

Evaluating measures of mental health attitudes

Survey Flow

EmbeddedData
PROLIFIC_PIDValue will be set from Panel or URL.

Standard: Information_Sheet (1 Question)
Standard: Consent (1 Question)

Branch: New Branch
If
If I have read and understood the information sheet for this study and consent to collection of my r... Yes Is Not Selected

EndSurvey: Advanced

Standard: Medworker (1 Question)

Branch: New Branch
If
If Are you currently working in a medical sector? Yes Is Not Selected

EndSurvey: Advanced

Standard: ProlificID (1 Question)
Standard: Age (1 Question)

Branch: New Branch
If
If What is your age? 17 or younger Is Selected

EndSurvey: Advanced

Block: Demographics (3 Questions)
Standard: RAQ (1 Question)
Standard: OMS_HC (2 Questions)
Standard: ATN_CHK (1 Question)
Standard: Links (1 Question)

EndSurvey: Advanced

Page Break

Start of Block: Information_Sheet

Information_Sheet

Evaluating Measures of Attitudes to Mental Health Recovery

Information Sheet

Dear Participant,

Hello, my name is Esario IV Daguman and I am completing a Master of Science in the School of Psychology at Massey University in New Zealand. I am researching how we measure attitudes towards mental health recovery and stigma. I would like to invite you to take part in this research.

What is the aim of the study?

The study aims to investigate two measures of attitudes about mental health recovery and stigma. These measures are used in recovery and stigma research, so we need to know how well they work and whether we can rely on them in ongoing research.

Who is eligible to participate?

I am inviting people registered on Prolific Academic who are at least 18 years old and working in the 'Medicine' employment sector to participate in the study.

What will participants be asked to do?

Should you agree to participate in this study, you will be asked to complete a brief questionnaire about mental health attitudes that includes 22 questions. The questionnaire will take about 10 minutes to complete. The survey is confidential. The questionnaire will begin when you click on the link and follow the steps to the questionnaire. You can skip questions if you like. It is not expected that participation will result in any discomfort.

What data or information will be collected, and what use will be made of it?

Age, gender, and ethnicity information will be collected in addition to your responses to questions from the two measures. All data is being collected only for this research. Initially, only my supervisor and I will have access to the data. The data will be stored on a password protected device only accessible by my supervisor and I. After we have analysed the data, we will ensure that any information in the dataset that could be used to identify you has been removed (for example, your Prolific ID number). Then we will publish the data in an online repository (on the Open Science Framework). The data contained in this repository will be accessible to other researchers and the public. Any personally-identifying information you provide (for example, your Prolific ID number) will be erased following the project's completion (July 2022). This data will be maintained in perpetuity after it has been anonymised.

How much will the participant be paid in exchange for their participation?

We will be compensating each participant for their time in taking part in the study with £1.25 (GBP).

What are your rights as a participant?

You are under no obligation to accept this invitation. If you decide to participate, you have the right to decline to respond to any question, ask any questions about the study at any time during participation, provide information on the understanding that your name will not be used, and be given access to a summary of the project findings when it is concluded. You may withdraw your data from the study up to two weeks after completing the questionnaire, without consequence. The completion and submission of the questionnaire implies consent.

What if I have any questions?

Please feel free to contact the researchers if you have any questions concerning the study. The contact details for the researchers are:

Esario IV Daguman
School of Psychology
Massey University, Auckland
Email: Esau.Daguman.1@uni.massey.ac.nz

Dr Joanne Taylor
School of Psychology
Massey University, Palmerston North
Phone: +64 (06) 9518068
Email: j.e.taylor@massey.ac.nz

Massey University Human Ethics Committee Statement

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email humanethics@massey.ac.nz

End of Block: Information_Sheet

Start of Block: Consent

Consent I have read and understood the information sheet for this study and consent to collection of my responses.

- Yes (1)
- No (2)

End of Block: Consent

Start of Block: Medworker

Medworker Are you currently working in a medical sector?

- Yes (1)
- No (2)

End of Block: Medworker

Start of Block: ProlificID

ProlificID Please enter your Prolific ID here.

_____

End of Block: ProlificID

Start of Block: Age

Age What is your age?

- 17 or younger (1)
- 18-20 (7)
- 21-30 (2)
- 31-40 (3)
- 41-50 (4)
- 51-60 (5)
- 61 or older (6)

End of Block: Age

Start of Block: Demographics

Gender What is your gender?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (5)
- Other (4) _____

Page Break

X→

Ethnicity What is your ethnicity?

- African (1)
- Black / African American (2)
- Black / British (3)
- Caribbean (4)
- East Asian (5)
- European (6)
- Latino / Hispanic (7)
- Māori (8)
- Middle Eastern (9)
- Mixed (10)
- Native American or Alaskan Native (11)
- Pacific Islander (12)
- Romani/Traveller (13)
- South Asian (14)
- SouthEast Asian (15)
- White / Sephardic Jew (16)
- White Mexican (17)
- Others (18) _____

RECOVERY ATTITUDES QUESTIONNAIRE (RAQ-7)
(Borkin et al., 2000)

Recovery is a process and experience that we all share. People face the challenge of recovery when they experience the crises of life, such as the death of a loved one, divorce, physical disabilities, and serious mental illnesses. Successful recovery does not change the fact that the experience has occurred, that the effects are still present, and that one's life has changed forever. Rather, successful recovery means that the person has changed, and that the meaning of these events to the person has also changed. They are no longer the primary focus of the person's life (Anthony, 1993).

We are interested in measuring your beliefs about the concept of recovery from mental illnesses.
Please read each of the following statements and using the scale below mark the rating that most closely matches your opinion.

	Strongly Agree (5)	Agree (4)	Neutral (3)	Disagree (2)	Strongly Disagree (1)
1. People in recovery sometimes have setbacks. (RD1)	○	○	○	○	○
2. To recover requires faith. (RP2)	○	○	○	○	○
3. Stigma associated with mental illness can slow down the recovery process. (RD3)	○	○	○	○	○
4. Recovery can occur even if symptoms of mental illness are present. (RP4)	○	○	○	○	○
5. Recovering from mental illness is possible no matter what you think may cause it. (RP5)	○	○	○	○	○
6. All people with serious mental illnesses can strive for recovery. (RP6)	○	○	○	○	○
7. People differ in the way they recover from a mental illness. (RD7)	○	○	○	○	○

End of Block: RAQ

Start of Block: OMS_HC

Opening Minds Scale for Health Care Providers (OMS-HC-15)
(Kassam et al., 2012; Modgill et al., 2014)

These questions ask you to agree or disagree with a series of statements about mental illness. There is no correct answer. Please mark the circle that best fits your opinion.

	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)
1. I am more comfortable helping a person who has a physical illness than I am helping a person who has a mental illness. (OA1)	○	○	○	○	○
2. If a colleague with whom I work told me they had a managed mental illness, I would be as willing to work with him/her. (OS3)	○	○	○	○	○
3. If I were under treatment for a mental illness I would not disclose this to any of my colleagues. (OD4)	○	○	○	○	○
4. I would see myself as weak if I had a mental illness and could not fix it myself. (OD6)	○	○	○	○	○
5. I would be reluctant to seek help if I had a mental illness. (OD7)	○	○	○	○	○
6. Employers should hire a person with a managed mental illness if he/she is the best person for the job. (OS8)	○	○	○	○	○
7. I would still go to a physician if I knew that the physician had been treated for a mental illness. (OS9)	○	○	○	○	○

Page Break

OMS_HC_2

Opening Minds Scale for Health Care Providers (OMS-HC-15)
(Kassam et al., 2012; Modgill et al., 2014)

These questions ask you to agree or disagree with a series of statements about mental illness. There is no correct answer. Please mark the circle that best fits your opinion.

	Strongly Disagree (1)	Disagree (2)	Neither Agree nor Disagree (3)	Agree (4)	Strongly Agree (5)
8. If I had a mental illness, I would tell my friends. (OD10)	○	○	○	○	○
9. Despite my professional beliefs, I have negative reactions towards people who have mental illness. (OA12)	○	○	○	○	○
10. There is little I can do to help people with mental illness. (OA13)	○	○	○	○	○
11. More than half of people with mental illness don't try hard enough to get better. (OA14)	○	○	○	○	○
12. I would not want a person with mental illness, even if it were appropriately managed, to work with children. (OS17)	○	○	○	○	○
13. Health care providers do not need to be advocates for people with mental illness. (OA18)	○	○	○	○	○
14. I would not mind if a person with a mental illness lived next door to me. (OS19)	○	○	○	○	○
15. I struggle to feel compassion for a person with a mental illness. (OA20)	○	○	○	○	○

End of Block: OMS_HC

Start of Block: ATN_CHK

ATN_CHK Please click on the '**Agree**' choice below to show that you are giving close and thoughtful attention to the question.

- Strongly Disagree (1)
- Disagree (2)
- Neither Agree nor Disagree (3)
- Agree (4)
- Strongly Agree (5)

End of Block: ATN_CHK

Start of Block: Links

Links
We thank you for your time in participating in the research.

The summary of the findings will be ready in approximately July 2022 and will be posted on https://psych-research.massey.ac.nz/

If you need help, information for psychological support resources is available on https://www.helpguide.org

Please click the next button to end the survey and be redirected to the Prolific app.

End of Block: Links

Appendix B – Acknowledgment of Low-Risk Ethics Notification

From: humanethics@massey.ac.nz
Subject: [HE007] - Human Ethics Notification - 4000025270
Date: November 23, 2021 at 9:56 AM
To: Esau.Daguman.1@uni.massey.ac.nz, J.E.Taylor@massey.ac.nz
Cc: humanethics@massey.ac.nz

Kia ora,

Link to the application
HoU Review Group

Ethics Notification Number: 4000025270
Title: Validation of the Recovery Attitudes Questionnaire and the Opening Minds Scale for Health Care Providers

Thank you for your notification which you have assessed as low risk.

Your project has been recorded in our database for inclusion in the Annual Report of the Massey University Human Ethics Committee.

The low risk notification for this project is valid for a maximum of three years.

Please notify me if situations subsequently occur which cause you to reconsider your initial ethical analysis that it is safe to proceed without approval by one of the University's Human Ethics Committees.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice-Chancellor and be in accordance with the Policy and Procedures for Course-Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

A reminder to include the following statement on all public documents:

"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named in this document are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you want to raise with someone other than the researcher(s), please contact Professor Craig Johnson, Director (Research Ethics), email humanethics@massey.ac.nz."

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish require evidence of committee approval (with an approval number), you will have to complete the application form again answering yes to the publication question to provide more information to go before one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

You are reminded that staff researchers and supervisors are fully responsible for ensuring that the information in the low risk notification has met the requirements and guidelines for submission of a low risk notification.

If you wish to print an official copy of this letter:

1. Please login to the RIMS system (https://rme.massey.ac.nz).
2. In the Ethics menu, select Ethics Applications.
3. Using the Advanced search with appropriate criteria to find only this application.
4. With the application on the Results tab, select Reports from the toolbar.
5. Select the "Human Ethics - Low Risk Notification Letter" link, this will open the report viewer.
6. Select the application code from the Report Parameters dropdown and submit. You can then select an export option from the top toolbar (Print, Save).

Yours sincerely
Professor Craig Johnson
Chair, Human Ethics Chairs' Committee and
Director (Research Ethics)

Appendix C – Detailed Participants' Demographic Information

Variable	*N*	%
Age		
17 or younger	-	-
18-20	116	40.56
21-30	97	33.92
31-40	40	13.99
41-50	19	6.64
51-60	8	2.80
61 or older	6	2.10
Gender		
Male	82	28.67
Female	202	70.63
Non-binary/ Third gender	2	0.70
Prefer not to say	-	-
Others	-	-
Ethnicity		
African	28	9.79
Black/ African American	2	0.70
Black/ British	7	2.45
Caribbean	1	0.35
East Asian	2	0.70
European	181	63.29
Latino/Hispanic	1	0.35
Māori		-
Middle Eastern	2	0.70
Mixed	7	2.45
Native American or Alaskan Native		-
Pacific Islander		-
Romani/Traveller		-
South Asian	13	4.55
SouthEast Asian	6	2.10
White / Sephardic Jew	33	11.54
White Mexican		-
Others		-
Unknown	3	1.05