A novel approach to investigate depression symptoms in the ageing population using Generalizability theory

A thesis

Submitted in fulfilment

of the requirements for the degree

of

Master of Science (Research) in Psychology

at

The University of Waikato

by

SRUTHY SAIJI KUMAR

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2022

Abstract

As depression is common in older people and confers a significant risk for dementia, its accurate assessment is essential to monitor and treat the condition. The 15-item Geriatric Depression Scale (GDS-15) is a widely used assessment tool for measuring depression in aged populations, and its psychometric properties have been recently improved using Rasch analysis. However, its temporal reliability and ability to distinguish between dynamic and enduring symptoms of depression have not been examined using the appropriate methodology. Distinguishing enduring aspects of depression helps to estimate risks of depression and long-lasting effects of an intervention while identifying and targeting dynamic symptoms may enhance efficiency of a treatment. Generalizability theory (G-theory) is a suitable method to distinguish between enduring and dynamic symptoms of depression, evaluate the overall reliability of the GDS-15, and identify sources of measurement error. This study applied G-theory to the longitudinal GDS-15 data of 354 participants aged 70 years and older from the Sydney Memory and Ageing Study, collected over ten years with 2-4 years intervals. The GDS-15 demonstrated strong reliability and generalizability of its scores across the sample population and assessment occasions in measuring enduring symptoms of depression (Ga= 0.82, Gr=0.90). In addition, three dynamic symptoms of depression were identified, namely helplessness, hopelessness, and loss of interest in activities, which did not affect the overall strong reliability of the GDS-15. This study demonstrated that the GDS-15 is a reliable measure for assessing enduring symptoms of depression and can be used to evaluate the efficacy of depression treatments and monitor depression levels in older adults. Dynamic symptoms identified by this study are more amendable and can be targeted on the first place to enhance effectiveness of a treatment.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to supervisor Dr. Oleg Medvedev for guiding me to complete this project. Thank you for the continuous feedback and encouragement to do my best.

I would also like to express my heartfelt gratitude and love to my husband Jithin and son Ethan for their unconditional love, support, and motivation. You both helped me to push boundaries and encouraged to fulfil my dreams.

Thank you, Mom, Dad, and Smruthy for continuous motivation and empathy. I am also grateful to my in-laws who supported me throughout my studies. My family always believed that I should never stop learning and persuaded to continue with my goal. A big thanks to all.

I would like to give special thanks to external advisors Dr Katya Numbers, Dr Alexander Merkin, Prof Perminder Sachdev, Prof Henry Brodaty, Dr Nicole Kochan, Prof Julian Trollor, and Dr Susan Mahon for providing me the data and helping with the reviewing my manuscript for journal submission. I also thank the MAS research team and the participants.

Co-Authored Works

The study presented in my thesis was submitted for publication as a co-authored work with my supervisor Dr Oleg Medvedev and the external advisory team working with the MAS data, including Dr Katya Numbers, Dr Alexander Merkin, Prof Perminder Sachdev, Prof Henry Brodaty, Dr Nicole Kochan, Prof Julian Trollor, and Dr Susan Mahon. The paper was published in the APA journal *Psychological Assessment* (Impact Factor=5.1) on March 31, 2022 (see Appendix A). I am the first and leading author and have made a major contribution to this study; however, this study was only successful with the guidance and support of my supervisor Dr Oleg Medvedev and other collaborators. My major contributions included designing the study, conducting statistical analyses, and writing and editing the manuscript that provided the foundation for this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1 Depression in older adults

Defining Depression and Late-life Depression

Depression is one of the most prevalent mental health issues in older adults worldwide (Blazer, 2003). Existing research has recognised depression as the third most prominent contributor to the worldwide burden of diseases (Collins et al., 2011). Depressive Disorders in the Diagnostic and Statistical Manual of Mental Disorders (5th ed; DSM-5) include major depressive disorder (MDD), persistent depressive disorder (also known as dysthymia), premenstrual dysphoric disorder, substance-induced specified depressive disorder, disruptive mood dysregulation disorder, other specified and unspecified depressive disorders. The common characteristics of depressive disorders involve sadness, feeling empty, irritability, cognitive and somatic symptoms that affect the significant functioning of the individual (American Psychiatric Association [APA], 2013). According to the DSM-5 criteria for making a diagnosis of depression, an individual should be experiencing five or more symptoms during the same two -weeks period and must at least have either a depressed mood or loss of interest or pleasure (APA, 2013). The symptoms include the following:

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (Note: In children and adolescents, can be irritable mood.)

2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).

3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)

4. Insomnia or hypersomnia nearly every day.

5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).

6. Fatigue or loss of energy nearly every day.

7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).

8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).

9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide. (APA, 2013, p.160)

The World Health Organization (WHO) identified depression as a major cause of disability worldwide, and it occurs among people of all age groups. The WHO considers age 65 years and above as older adults concerning the individual's competency in accommodating environmental factors (Sözeri-Varma, 2012). Most studies on late-life depression have used the definition of Major Depressive Disorder with onset at the age of 60 or above (Naismith et al., 2012; Sözeri-Varma, 2012). Espinoza and Kaufman (2014) defined late-life depression (LLD)

as depression occurring after 60 years. Depressive symptoms that occur for two weeks but do not meet all the criteria for MDD, called sub-threshold or subsyndromal or minor depressive disorder, are more frequently reported among the older population (Blazer, 2009).

A considerable amount of literature has also distinguished between early-onset depression, which has onset before the age of 60 with recurrent depressive episodes in late-life and late-onset depression that develops after the age of 60 (Alexopoulos, 2005; Disabato et al., 2014; Janssen et al., 2007; Sachs-Ericsson et al., 2013). In addition, these studies have identified differences concerning symptom presentations, neurobiological factors, course, and prognosis of depression. For example, Rapp et al. (2005) distinguished between early-onset and late-onset depression based on neuropsychological functioning, depressive symptoms and co-morbidity. They found that older adults with late-onset depression exhibited deficits in executive functioning and attention, increased rates of anhedonia and higher co-morbidity to cardiovascular illnesses than older adults with early-onset depression.

Moreover, another longitudinal study that compared differences in early and late-onset depression symptoms in the older population found higher suicidal ideation and suicidal behaviour in individuals with early-onset depression than those with late-onset depression (Sachs-Ericsson et al., 2013). There are also gender differences in symptoms of depression in the older population. For instance, men show anger, apathy, anhedonia but not sadness, whereas women show somatic symptoms and dysphoria (Espinoza & Kaufman, 2014). Studies have identified several challenges in assessing depression among older adults, and their results indicated that a better understanding of depression symptoms in older

adults is vital. Accurate measurement of depression is fundamental to identify and assess the risks of developing depressive disorders and intervene at early stages. Some depression symptoms may be dynamic, and identifying these symptoms is essential as it could be a primary target of an intervention because they are more amenable to change, potentially enhancing treatment efficiency.

Prevalence and Co-morbidity of Late-life Depression

Epidemiological studies on depression indicate that the prevalence of late-life depression is substantial, affecting 1-7% of the general older adult population and up to 13.5% of older individuals who receive home healthcare (CDC, 2017; Gonçalves-Pereira et al., 2019; Kennedy, 2015; Reynolds et al., 2015). According to systematic reviews and meta-analyses on MDD, worldwide prevalence rates of MDD range between 1% to 5% among older adults aged 65 and older (Blazer, 2003; Blazer, 2009; Djernes, 2006; Fiske et al., 2009). The global prevalence estimate of sub-threshold depression, which does not meet the complete criteria for MDD, is around 15 % among older adults aged 65 years or above (Blazer, 2009; Fiske et al., 2009; Haigh et al., 2018). However, the prevalence rate of depression varies noticeably based on the sample studied and may depend on diagnostic accuracy.

For example, a community-based study conducted in Brazil on the older population aged 75 years and older found an 11.1% prevalence of late-life depression and 25.6% prevalence of clinical depressive symptoms (Leles da Costa Dias et al., 2019). On the other hand, a study completed among the older adults residing in nursing homes in Mexico found that almost 40% of the older adults presented depressive symptoms (Arias-Merino et al., 2012). The distinction between dynamic and enduring depression symptoms can play an essential role in

improving the accuracy of estimating prevalence rates of depression. It would also reduce over-diagnosis of depression as dynamic symptoms often subside shortly after an assessment.

Furthermore, studies comparing the prevalence of depression among older adults in rural and urban areas reported higher rates in urban areas compared to rural areas (Mohd Sidik et al., 2003; Schulman et al., 2002). A substantial amount of research indicated that people residing in urban areas are at more risk of developing mental health illnesses compared to those in rural areas (Galea et al., 2011; Lambert et al., 2015; Srivastava, 2009). Similar to the gender ratio among younger adults, research indicated that older women are at a higher risk of depression compared to men (Cole & Dendukuri, 2003; Djernes, 2006; Luppa et al., 2012). However, Bebbington (1996) pointed out that these prevalence differences in depressive symptoms between men and women are mainly due to social factors such as role, support, life events and psychological aspects such as coping style rather than genetic vulnerability. The study by Bebbington (1996) seemed to suggest that some of the depression symptoms could be dynamic due to environmental and social factors, whereas others may be enduring that could be attributable to genetic factors. This raises the need for a clear distinction between enduring and dynamic symptoms using appropriate methodology.

Studies have also attempted to distinguish depression in older adults based on their first encounter- older adults experience the first depression episode in old age (late-onset) versus the first episode of depression before old age (early-onset). For example, a study conducted by Brodaty et al. (2001) compared early-onset and late-onset depression in older adults attending the Mood Disorders Unit and found that more than 50 per cent of the reported cases of depression had their first

episode at the age of 60 or above. Another epidemiological study on depression in the older population identified the prevalence of subthreshold depression as two times higher than major depressive disorder (Meeks et al., 2011). Whilst studies have been carried out to distinguish between late-onset and early-onset depression; there is limited evidence that considers dynamic and enduring symptoms in measuring depression. This might impact the reliability of the measurement of depression as dynamic symptoms often subside after an assessment, and a lack of reliable distinction between dynamic and enduring symptoms could potentially lead to misdiagnosis.

The prevalence of late-life depression varies across different countries due to different reasons. Studies have identified variability in the quality of life, mortality rate, age, mean income, lack of appropriate universal screening tools, attribution of depressive symptoms to physical ailments, samples studied and diagnostic methods that may account for differences across countries. (Arias-Merino et al., 2012; Sims et al., 2012). For instance, prevalence rates of depression range from 2.4% in Japan and 2.7 % in the US up to 35% in China and even 64.3 % in Mexico (Forlani et al., 2014; Gonçalves-Pereira et al., 2019; Haigh et al., 2018; Leles da Costa Dias et al., 2019). Another issue that may impact varying late-life depression prevalence estimates across epidemiological studies is the lack of tools reliably distinguishing dynamic and enduring symptoms of depression.

Challenges for accurate diagnosis of depression in older adults are also related to variability in the clinical presentation of the disease in older age and overlap of the symptoms with other mental conditions, including delirium (Grayson & Thomas, 2013; Haigh et al., 2018). In addition, older adults with

depression tend to focus more on somatic symptoms instead of reporting dysphoria or sadness, which often lead to its underdiagnoses (Haigh et al., 2018). Factors affecting the recognition of depression symptoms have been investigated in various studies suggesting that older adults are less likely to express their moods (Sözeri-Varma, 2012). They also often attribute depression symptoms such as loss of interest and anhedonia to normal ageing process (Jorm, 2001; Kaya, 1999). Additionally, research to date has not yet distinguished dynamic (state) and enduring (trait) symptoms of depression reflected by specific items in different measures, which again impacts on the accurate psychometric measurement of depression in the older population (Kumar et al., 2022). As research on prevalence rates of depression is conducted either directly or indirectly through online or using an app, a state-trait distinction and an accurate psychometric measurement are essential to reduce false positives and false negatives dependent on the assessment occasion.

A considerable amount of research on depression in older adults indicated that the clinical presentation, causal factors, and treatment responses are different in older adults compared with young adults (Kaya, 1999; Sözeri-Varma, 2012). There are differences in depression symptoms among young and older adults; however, Haigh et al. (2018) note that empirical evidence is limited to generalizable conclusions. Furthermore, there is limited literature on whether these differences result from physiological changes (Sözeri-Varma, 2012). Therefore, a more accurate assessment of depression and distinction between dynamic and enduring symptoms will help to identify risks and intervene to reduce devastating consequences of this condition in the ageing population. Further research is needed to refine the available psychometric assessments using

appropriate methodologies and identify sources of measurement errors while measuring depression.

Etiology and Impact of Late-life Depression

Researchers have identified late-life depression as a public health priority as it can have devastating consequences among the older population (Haigh et al., 2018; Meeks et al., 2011). Recent research suggests that depression can be a feature of the prodrome or early stages of dementia (Baune et al., 2018; Blazer, 2003; Fang et al., 2018; Haigh et al., 2018; Livingston et al., 2020; Naushad et al., 2018; Perera et al., 2017; Read et al., 2017; Teo et al., 2018; Wang et al., 2018). Extensive research has shown that depression in older adults is linked to increased co-morbidity with other disorders, decreased social and cognitive functioning, self-neglect and increased mortality (Blazer, 2003; Naushad et al., 2018; Perera et al., 2017; Read et al., 2017; Teo et al., 2018; Wang et al., 2018). These studies collectively imply the critical role of accurate psychometric measures for assessing depression in older adults, which is essential for identifying and monitoring depression at earlier stages and providing an appropriate treatment to reduce its consequences.

Biological Factors

A growing body of research indicates that depression symptoms in the older age group are related to several biological, social, psychological, behavioural, and demographic factors. A recent systematic review and meta-analysis of gene studies of late-life depression suggested a significant genetic component in its aetiology, indicating that genes are associated with hippocampal plasticity and stress reactivity (Tsang et al., 2017). Furthermore, they found that the dysregulation of these pathways contributes to depression in late life. Twin

studies on LLD have identified low to moderate heritability, with genetic factors explaining 14% to 55% of the variance, while the remaining variance explained non-shared environment influences (Carmelli et al., 2000; Johnson et al., 2002). The environmental influences may cause dynamic or state changes, whereas the genetic factors are more related to enduring or trait aspects of depression, which supports the notion that a clear distinction between state and trait aspects of depression is crucial.

A comprehensive study on biological risk factors of LLD identified physiological illnesses as leading causal factors to LLD; following by vascular diseases such as myocardial infarction, coronary heart disease, hyperhomocysteinemia, and general health factors such as obesity, diabetes mellitus, poor health status and poor self-perceived health (Aziz & Steffens, 2013). Alexopoulos et al. (1997) coined the term vascular depression hypothesis for the known co-morbidity of vascular diseases with depression and the relation between ischemic lesions and distinct behavioural symptoms. It suggested that "the disruption of prefrontal systems or their modulating pathways by single lesions or an accumulation of lesions exceeding a threshold are hypothesised to be central mechanisms in vascular depression" (Alexopoulos et al., 1997). Such patients show more psychomotor retardation and less psychomotor agitation, less guilt, poor insight, and more significant cognitive impairment than those without nonvascular depression (Aziz & Steffens, 2013). Much of the previous population and imaging studies suggests that cerebrovascular injury and white matter hyperintensities are associated with LLD. The higher homocysteine levels lead to cerebral vascular diseases and deficiency of neurotransmitters, increasing risks for LLD (Culang-Reinlieb et al., 2011; Folstein et al., 2007).

Furthermore, neuroimaging studies have identified several areas of the brain, including the anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), and the hippocampus, to be associated with LLD (Benjamin & Steffens, 2011; Taylor et al., 2007). A meta-analysis of the magnetic resonance imaging studies on brain volume anomalies found that individuals with depression had smaller ACC volumes (Koolschijn et al., 2009). Studies in LLD have also demonstrated reduced hippocampal volume and smaller OFC volumes in depressed older adults than non-depressed older adults (Benjamin & Steffens, 2011; Koolschijn et al., 2009). Additionally, medical co-morbidities such as myocardial infarction, coronary heart disease, obesity, diabetes, and neurological disorders such as Parkinson disease can also lead to depression in older adults (Aziz & Steffens, 2013). Overall, these studies highlight the pertinent role of reliable assessment of depression because it ensures reliable and valid empirical data for identifying depression symptoms and linking them to specific brain pathologies.

Psychosocial Factors

1. Stressful life events

Psychological and social causal factors of depression are often interconnected. Most of the studies on aetiology and psychological correlates of depression to date have been focused mainly on adults and young adults. Although the underpinning psychological causal factors of the LLD are similar to those in adulthood, studies indicate that stress is one of the main contributing factors of depression. Studies proposed that stressful life events make significant long-lasting changes in the brain's biology by affecting the functioning of neurotransmitters and intraneuronal signalling systems, making a person more susceptible to subsequent episodes of depression even without external stressors

(Sadock et al., 2015). Recent stressful events are a strong predictor of the onset of Major Depressive Episode (MDE). However, there have been mixed opinions on the impact of stress on the onset of depression, with some clinicians arguing that life stressors have a partial role in the onset of depression (Sadock et al., 2015). Stressful life events are likely to be associated with dynamic symptoms reflecting state aspects. They may be temporary, thereby influencing the accurate assessment of depression as some of the symptoms may either diminish after an assessment or may be affected by the assessment occasion.

2. Personality factors

No personality trait uniquely influences a person to develop depression, even though certain personality traits or personality disorders may pose a greater risk for MDD and LDD. Numerous studies have considered personality factors, especially neuroticism, low extroversion, as an essential predictor of the onset of LLD (Hayward et al., 2013; Steunenberg et al., 2010). It has been suggested that individuals with a high score on neuroticism tend to respond inadequately to stress and interpret situations as intimidating, difficult or hopeless (Aziz & Steffens, 2013). Although neuroticism is a personality trait, it is also associated with mood changes, a state aspect that can lead to misdiagnosis of depression if there is no clear differentiation between dynamic or enduring symptoms.

3. Psychodynamic theories of Depression

Several studies have identified ambivalent emotional expression and insecure attachment leading to late-life depression (Morse & Lynch, 2004; Paradiso et al., 2012). One of the most classic views of depression is the psychodynamic theory of depression, defined by Sigmund Freud and extended by Karl Abraham. It proposes that: disruptions in the infant-mother attachment relationship during the

oral stage can predispose a person to develop depression in later life; depression is related to real or imaginary object loss; introjection of the lost objects is a defense mechanism used to deal with the distress associated with the object's loss; and because the lost object is involved love and hate, feelings of anger are directed towards self (Sadock et al., 2015, p. 354). In addition, the attachment theory proposed by John Bowlby suggests that problems in early attachments and traumatic separation in childhood is a risk factor to the onset of depression as any loss in adulthood would recuperate the traumatic loss in childhood and acts as a precipitating factor for depression (Bowlby, 2008; Sadock et al., 2015). However, there is a lack of empirical evidence supporting these theories, perhaps due to measurement issues.

4. Cognitive theory of Depression

Another well-known theory regarding depression is the cognitive theory, developed by Aaron T. Beck, which posits that depression results from cognitive distortions present in an individual vulnerable to depression (Beck, 1967). It proposed that depression may be caused by the tendency to infer day-to-day events negatively. These types of thoughts are called cognitive errors, and after several cycles of adverse events in childhood, the individual may develop a deep-rooted negative schema, an enduring negative cognitive belief system about some aspects of life. These are cognitive templates that distinguish internal and external data in ways reformed by early life experiences, which are often referred to as depressogenic schemata. According to Beck's cognitive triad of depression, the maladaptive schematic representation of the self, world and future are triggered by corresponding life experiences. It leads to a preferential processing bias and the development of negative automatic thoughts and interpretations. A negative

attitude results in a negative self-concept, which becomes pathogenic when it leads to a negative value judgement (Beck et al., 2014).

According to Beck and Alford (2009), individuals with depression express negative feelings about themselves caused by negative attitudes about themselves (such as 'I am worthless') directed towards self. Furthermore, the interpretation of their experiences, description of its occurrence and views about the future reflects personal deficiency, self-blame, and negative expectations (pessimistic view) of the future (Beck et al., 2014). The cognitive theory proposes that when these components get activated, individuals interpret their experiences as a personal defeat; they attribute the defeat as caused by a defect in themselves; regard themselves as worthless; blames themselves for that attribute and dislike themselves; they perceive the future as hopeless or filled with pain (Beck et al., 2014). Consequently, they experience sadness, lack of motivation, thoughts of self-harm and retardation or agitation. While cognitive theorists proposed the idea of schemas leading to depression, symptoms such as the feeling of sadness, lack of motivation and worthlessness may also appear because of temporary changes in mood or external events rather than due to schemas associated with enduring patterns in depression. Therefore, it is necessary to identify dynamic and enduring components of depression as some symptoms can be dynamic while others can be long-lasting.

5. Learned helplessness theory of Depression

The learned helplessness theory of depression was introduced by psychologists Martin Seligman and Steven F. Maier. It was first experimented with helpless behaviour in a laboratory when the dogs were subjected to electrical shocks from which they could not escape. In a reformulated view of learned

helplessness into human depression, internal causative explanations lead to a loss of self-esteem after adverse external experiences. It is based on the idea that an individual experiences depression when they attribute the events as uncontrollable. According to this view, depression results from the real or perceived absence of a sense of control and mastery of the environment (Maier & Seligman, 1976; Sadock et al., 2015; Seligman, 1972).

Seligman (1972) pointed out that anxiety comes as the first response to a stressful event and depression results from marked hopelessness about coping with stressful events. The negative attributional style of depression is (1) internal, that the person attributes adverse events as their failure, (2) stable, where an individual may believe more bad things will happen, even after a specific adverse event had passed; and (3) global, attributions range across a variety of matters (Barlow, 2002; Barlow et al., 2016). Studies indicate that negative attributional style itself did not predict the development of depression; instead, they act as a precipitant and are a risk factor for depression (Abramson & Alloy, 2006). A modified version of the learned helplessness theory by Abramson et al. (1989) deemphasized the impact of negative attributional style and highlighted the sense of hopelessness as a significant causal factor for depression. According to the hopelessness theory, an individual gives up and becomes hopeless about regaining control over adverse events, which causes depression. Although hopelessness can act as a precipitating factor in depression, people often feel helpless and hopeless due to current life events. The feelings of helplessness and hopelessness may also be dynamic and often recede after some time, which may or may not lead to learned helplessness and depression. Therefore, distinguishing a person's feelings of hopelessness and helplessness as a dynamic or enduring component would be beneficial to formulate successful interventions.

Conclusion

Depression is a prevalent mental health issue causing disability in all age groups and countries. Late-life depression is one of the common mental health illnesses diagnosed in the older population. However, epidemiological studies suggest that LLD often goes underdiagnosed, mainly due to problems with limited resources for an accurate assessment. The causal factors of LLD are quite different from depression in adulthood. Unlike MDD in adults and young adults, it is evident that physiological illnesses, neurological and biological factors play a significant role in the onset of LLD. However, the literature on LLD indicates limited empirical evidence on various psychological factors and correlates of LLD and its impact on older adults. Reliable and valid research is the core for a better understanding of LLD. More research into LLD is necessary to understand LLD better, and to develop effective prevention and intervention strategies. The studies mentioned above provide essential insights about LLD and factors influencing its accurate measurement. Reliable measurement and distinction between enduring and dynamic symptoms of depression would be crucial for accurate assessment and to reduce misdiagnosis of LLD. Therefore, investigating the reliability and generalisability of assessment scores when measuring depression is vital for both research and clinical practice.

Chapter 2 Measuring Depression in Older Adults

Overview of Depression Assessment Tools

In addition to the DSM-5 criteria for diagnosis of depressive disorders, there are numerous assessment tools designed for assessing depression. Although some instruments measuring depression among adults have also been validated in adults and older populations, the Geriatric Depression Scale-15 (GDS-15; Sheikh & Yesavage, 1986) is the only instrument specifically designed to assess depression in older adults. The other commonly used self-report measures of depression involve the Center for Epidemiologic Studies Depression Scale (CES-D; Sawyer, 1977), Zung Self-rating Depression Scale (SDS; Zung, 1965), Beck Depression Inventory (BDI; Beck et al., 1996), and the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983). Table 1 lists the psychometric properties of common depression measures that have over 1000 citations in Google Scholar (Agrell & Dehlin, 1989; Bjelland et al., 2002; Dunn & Sacco, 1989; Gabrys & Peters, 1985; Lopez et al., 2010; Snyder et al., 2000; Trajković et al., 2011; Yesavage et al., 1982; Zung et al., 1965). Among all the depression measures, GDS was the only measure specifically designed for older adults, and all other measures were developed to assess depression severity in the general population.

The Center for Epidemiologic Studies Depression Scale (CES-D) measures the current level of depression symptoms in a general population, and it has the highest number of citations (Table 1). It has been shortened for different populations, a 10-item and 5- item version for older adults, a 9-item version for screening patients with rheumatoid arthritis (Norris et al., 2004; Irwin et al., 1999; Martens et al., 2006). The 20 items in CES-D measure perceived mood and level of functioning during the last week. While CES-D measures four factors -

depressed affect, positive affect, somatic problems and retarded activity, and interpersonal relationship problems, emphasising depressed affect, GDS omits somatic symptoms as they may interfere with age-related symptoms. However, CES-D items do not assess the diagnostic criteria of appetite, anhedonia, psychomotor agitation or retardation, guilt, or suicidal ideation (Smarr & Keefer, 2011). The CES-D is utilised as a screening tool to classify people at risk for clinical depression.

Another commonly used tool is the Hospital Anxiety and Depression scale (HADS) to assess anxiety and depression symptoms in the general medical population aged 16 and 65 years. It is a 4-point Likert scale instrument with 14 items scored from 0 to 3, and with seven depression items measuring cognitive and emotional aspects of depression, mainly anhedonia, and seven items measuring cognitive and emotional components of anxiety (Smarr & Keefer, 2011). Items measuring somatic symptoms associated with emotional and physical disorders are omitted. However, it was not developed for the older population.

The Beck Depression Inventory (BDI) measures depression symptoms and severity in people aged 13 years and above (Beck et al., 1961). The BDI-II is the commonly used version (Beck et al., 1996). The revision omitted items related to weight loss, body image, hypochondria, and working difficulty corresponds to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria (Smarr & Keefer, 2011). The BDI-II consists of items measuring cognitive, affective, somatic, and vegetative symptoms of depression (Smarr & Keefer, 2011). It is a self-report 4-point scale indicating the degree of severity in

which the items are rated not at all (0) to extreme form (3). When compared to BDI-II, GDS is easy to administer among the older population.

The Zung Self-rating Depression Scale (SDS) developed by Zung (1986) is a self-administered 20 item assessment tool for depression and mood. It is a Likert type scale with scores for each item ranging from 1 to 4, and the SDS score ranges between 20 to 80. The SDS appears to have low to moderate reliability of scores in older adults, possibly due to the somatic items on the scale (Fiske et al., 2009)

Table 1 *Psychometric properties of common depression assessment tools and the number of citations of each scale in Google Scholar (5 April 2022)*

Scale	Original reference	Target population	Items	Cronbach's alpha	Reliability (test-retest)	Validity	Google Scholar
GDS-30	Yesavage et al, 1982	Older adults	30	0.69-0.99	0.85-0.94	SDS *r*=0.84-0.88; CES-D *r*=0.82; BDI *r*=0.78	17115
GDS-15	Sheikh & Yesavage, 1986	Older adults	15	0.75-0.90	0.58	GDS *r*=0.60-0.91; MADRS *r*=0.78	7345
SDS	Zung, 1965	Adults	20	0.79-0.93	Not reported	GDS *r*=0.59-0.88; BDI *r*=0.85-0.86; CES-D *r*=0.81; MMPID *r*=0.59-0.79	12386
BDI-II	Beck et al., 1996	Adults	21	0.76-0.93	0.93	CES-D *r*=0.69; PSS *r*=0.64; GDS *r*=0.71	5530
CES-D 20	Sawyer, 1977	Adults	20	0.86-0.90	Not reported	BDI-II *r*=0.69-0.75; GDS *r*=0.82; SDS *r*=0.81	58431
HADS-D	Zigmond & Snaith, 1983	Medical adults	7	0.67-0.90	*ICC**= 0.69-0.88	MADRS *r*=0.62-0.81; BDI *r*=0.62-0.70; GDS *r*=0.72	44124

Notes. *Intraclass correlation coefficient; SDS= Zung Self-Rating Depression Scale; GDS= Geriatric Depression Scale; BDI-II= Beck Depression Inventory; CES-D= Center for Epidemiologic Studies Depression Scale; HADS-D= Hospital Anxiety and Depression Scale; PSS= Perceived Stress Scale; MMPI D= Depression subscale of Minnesota Multiphasic Personality Inventory; MADRS= Montgomery-Asberg Depression Rating Scale

The Geriatric Depression Scale

The Geriatric Depression Scale (GDS), developed by Sheikh & Yesavage (1986), is one of the widely used self-report instruments for assessing depression that was developed explicitly for older adults (Sjöberg et al., 2017; Tsoi et al., 2017; Yesavage et al., 1982). The GDS-15 recognises depression while differentiating it from dementia (Smarr & Keefer, 2011). The original version contains 30 items, while the shorter form has 15 items (GDS-15), evidence suggesting it is equally reliable (Pocklington et al., 2016). The GDS-15 is a screening tool for geriatric depression rather than diagnostic classification (Sheikh & Yesavage, 1986). The shorter version was developed to improve its acceptability, reduce fatigue or lack of focus in the older population (Smarr & Keefer, 2011). Its 15 items represent symptoms covering affective (e.g., sadness, apathy, crying) and cognitive (e.g., worthlessness, hopelessness, helplessness, guilt) domains of depression. Unlike other tools, the GDS-15 does not assess somatic symptoms such as pain, and physical symptoms may be inadvertently confused or conflated (Pocklington et al., 2016; Smarr & Keefer, 2011).

The GDS-15 is easy to administer, requires a shorter time to complete, and several studies support the reliability and validity of its test scores (Chiesi et al., 2017; Chiesi et al., 2018; Guerin et al., 2018; Niu et al., 2018; Sheikh & Yesavage, 1986). In addition, the GDS-15 demonstrates high internal consistency, with Cronbach's α ranging from 0.74 to 0.86 (Friedman et al., 2005; Marwijk et al., 1995). A current systematic review on the diagnostic accuracy of the GDS-15 found sensitivity and specificity of 0.89 and 0.77, respectively, for detecting clinical representation of depressive symptoms at a cut-off score of 5/6 (Benedetti et al., 2018; Pocklington et al., 2016). Further, the reliability and psychometric

properties of the GDS-15 were recently evaluated and improved using Rasch methodology, with transformation tables created to convert GDS-15 raw scores into interval-level data now available to researchers (Merkin et al., 2020).

Studies investigating psychometric properties of GDS-15 have reported good construct validity and criterion validity of test scores (Friedman et al., 2005). For example, a validation study of the GDS-15 on non-demented community-residing older adults found an overall sensitivity of 0.97 and specificity of 0.95 at a cut off score of 4 (Nyunt et al., 2009). Evidence also suggests that all the items of GDS-15 discriminate between individuals with depression and those without depression (Acousta Quiroz et al., 2021). Furthermore, another study reported reliability of scores and convergent validity of GDS-15 scores with other depression measures among immigrant Chinese in Australia (Dow et al., 2018). However, there are also reports of inconsistencies within the literature suggesting various factors in the GDS-15. For instance, Imai et al. (2014) identified three factors- energy loss and pessimistic outlook (factor 1), positive mental status (factor 2) and empty feeling (factor 3). In addition, they found correlations between factor 1 and quality of life, factor 2 with daily life activities and factor 3 with subjective cognitive functioning.

On the other hand, a meta-analysis by Kim et al. (2012) on the factor structures of different language versions of the GDS-15 found the English and Chinese versions as having four factors each, while the Korean version has six, and items making up these factors also vary by language. Thus, although psychometric properties of GDS-15 have been examined in various studies and have found GDS-15 as a valid and reliable tool for screening depression in community-dwelling older populations, evidence is limited to name the factors

equally (Quiroz et al., 2021). A more recent study by Merkin et al. (2020) achieved a good fit of the GDS-15 to the unidimensional Rasch model identical to the one-parameter Item Response Theory (IRT) model, which supported psychometric properties, unidimensionality and internal validity of the GDS scores and its items. Merkin et al. (2020) has also found that the GDS items were invariant across personal factors such as age, sex, and diagnosis (e.g., MCI). However, there is a paucity of studies examining the distinction between the dynamic and enduring symptoms of depression (Beijers et al., 2019).

State and Trait Distinction in Measuring Depression

Distinguishing dynamic versus enduring aspects of depression helps refine the test scores and their context-specific applicability and adds information about the sources of measurement errors when assessing affective disorders. Generally, traits refer to relatively enduring attributes of a person. It implies individual differences in vulnerability to the disorder and perception of a stimulus as depressive (Guillot-Valdés et al., 2020). In contrast, states refer to reactions dependent on the current context and can reflect an adjustment to a specific situation (Gana et al., 2016). There has been a trait versus state model of depression since the 20th century (Ban, 2014). The binary model of depression proposes two types of depression: endogenous depression and reactive depression (Bech et al., 1988; Mizushima et al., 2013). Endogenous depression denotes trait depression, which is stable against external influences, referring to enduring features likely mediated biologically. In contrast, reactive depression is referred to as state or dynamic depression, which is more dependent on current life situations and events (Bech et al., 1988; Gana et al., 2016).

People with depression experience emotional, psychological, and physiological symptoms of decreased mood, diminished interest in activities, low self-worth, fatigue, increased or decreased appetite, and difficulty concentrating that may change over time (Berman & Furst, 2011). Goldston et al. (2006) found traits but not state aspects of the Beck Depression Inventory were significantly linked to suicide attempts. However, there is a paucity of studies examining the distinction between the dynamic and enduring symptoms of depression (Beijers et al., 2019). Only a few assessments tools measure both state and trait depression, such as the Diagnostic Melancholia Scale (Bech et al., 1988) and Maryland Trait and State Depression Scale (Chiappelli et al., 2014).

The psychometric properties of these scales have not been rigorously investigated and hence, do not assure the generalizability of results. Therefore, identifying stable and dynamic depressive symptoms is essential for both research and clinical practice. Thus, there is a need to distinguish between enduring (state) and dynamic (trait) aspects of depression across widely used assessment tools of depression, such as the GDS-15, using appropriate methodology. Clinicians should seek ways to capture better dynamic (e.g., acute) vs enduring symptoms of depression, and researchers should better understand the tools that capture these differences. Moreover, assessing enduring symptoms of depression will permit a more comprehensive assessment to estimate risks for depression. Evaluation of dynamic aspects will inform the development of interventions targeting this affective condition's dynamic and most amenable symptoms.

Conclusion

In addition to the DSM-5 criteria for diagnosis of depression, there are several instruments available to assess the depression symptoms and severity in a

range of populations. However, only GDS and its shorter versions were the most common assessment tools used explicitly for measuring depression in older adults. The psychometric properties of the GDS-15 suggest it is a reliable and valid tool for assessing depression in ageing populations, and it is easy to administer. However, studies are limited in exploring the dynamic and enduring properties of depression assessed by the GDS-15 and its temporal stability. A distinction between dynamic and enduring symptoms within a measurement helps to reduce errors in assessment and misdiagnosis as well as to identify the most amenable symptoms. Therefore, differentiating dynamic and enduring symptoms of depression in GDS and different depression assessment tools need further exploration.

Chapter 3 Theories of measurement

Measurement refers to assigning scores to individuals such that these scores denote characteristics of the individuals (Nunnally, 1994). Everitt (2006) defined psychometrics as "the study of measuring psychological characteristics such as abilities, aptitudes, achievement, personality traits and knowledge" (p.3). Traditionally, psychometrics has been widely applied in intelligence and achievement testing. However, recent psychometrics works are mostly associated with measuring unobserved or latent traits and variables (Wilson & Gochyyev, 2014). The most important concepts within psychometrics are test reliability and test validity. The test reliability indicates how consistently a test measures what it is supposed to measure. A test is considered valid if it measures what it claims to be measuring (Wilson & Gochyyev, 2014). Test reliability is one of the most common concepts in Classical Test Theory (CTT) and Generalizability theory (G-theory). CTT is often described as the foundational measurement model (Cardinet et al., 2011). This chapter discusses and compares the core constructs in CTT and G-theory and explains how G-theory has been utilised in distinguishing state or dynamic and trait or enduring aspects in a measure.

Classical Test Theory

The classical test theory provides the foundations of the true-score model and the reliability estimations (De Gruijter & Leo, 2007). CTT postulates basic assumptions about relationships between the true, error, and observed scores. In the CTT model, an individual's observed score on a test is assumed to combine an individual's true score and random measurement error. Thus, CTT proposes that an observed score (X) on a test is the sum of a true score (T) and an error score (E) (Lord & Novick, 1968). It is defined by the equation as follows:

$$X = T + E \qquad (1)$$

The true score corresponds to an individual's expected observed score on an assumed infinite number of repeats of the measure with no carryover effects (Vispoel et al., 2018). Based on the assumption that both *T* and *E* are independent and variances are additive, the variance (σ) of the observed score is the total variance of the true score and the error score variance. It is represented as follows:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \qquad (2)$$

The reliability of the test is central to CTT, and it refers to the accuracy in measurement, described as the consistency of test scores over repeated measurements (Brennan, 2001). The assumptions of CTT by Lord & Novick (1968) lead to the formulation of the ratio of reliability coefficient, that is, the degree of true variance within the total observed variance, expressed by the formula as follows:

$$\text{Reliability coefficient} = P_{xx} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} \qquad (3)$$

where σ_T^2 *is True variance and* σ_E^2 *is total error variance.*

The computation of reliability coefficient in the CTT is based on the parallel test assumption that "parallel forms of a measure can be created in which an individual has the identical true score on both forms, variances of observed-score are same across forms and error scores are not correlated either with true score or each other (Vispoel et al., 2018). The standard error of measurement (SEM) is "the inconsistency within-person in score-scale units" (Wang & Osterlind, 2014, p.37). It denotes the standard deviation of an assumed set of

repeated measurements on the same person. The SEM is calculated after the reliability coefficient is computed, as follows:

$$SEM = \sigma_X \sqrt{1 - P_{XX}} \quad (4)$$

Equation 4 implies that SEM = 0 if there is only true variance. In the CTT, the reliability of test scores and error variance is restricted to a single element such as the test items (Cronbach's alpha), the occasion (test-retest), or the rater (inter-rater reliability). The internal consistency analyses investigate errors related to items' differences, while test-retest reliability evaluates errors associated with variations across time. Other sources of error are encompassed within the "true" score in both circumstances. For example, this variation could be attributable to the systematic error, the measuring item, or multiple testing occasions, but CTT cannot distinguish between these many causes of error. Even though the test-retest reliability coefficient is a standard CTT method for evaluating the consistency of assessment scores over time and differentiating measurements of state versus trait, this method does not consider multiple error sources such as test items, occasions, and their interactions with an individual (Medvedev et al., 2021).

Generalisability Theory

G-theory is an advanced statistical method that estimates all possible error variances and their influences on the overall reliability (Cronbach et al. 1963; Truong et al., 2020). It also emphasises how precisely the observed scores enable us to generalise the participants' behaviour in a defined set of situations (Shavelson et al., 1989). G-theory can be considered as an extension of Classical Test Theory (CTT), which postulates that every test score consists of true and error variance, but it transcends CTT’s assumption of considering error as a single factor (Cronbach et al. 1963). In natural environments, there are different sources

of influence such as individual's personality, characteristics of measurement tool, and occasional (e.g., time) that may contribute to measurement error. G-theory based studies allow researchers to investigate score reliability by simultaneously finding several systematic and unsystematic measurement errors (Brennan, 2010; Shavelson et al., 1989). G-theory is applied at the item level since the whole point of this method is to break down variance into its atomic or constituent parts, which means down to the item level (Forrest et al., 2021). In fact, the first papers on G-theory used binary outcome measures as item responses (1, 0) to demonstrate how G-theory worked and the appropriateness of binary data for this methodology (Brennan, 2001).

Unlike Classical Test Theory (CTT), which considers a single aspect of reliability such as test-retest or internal consistency, G-theory evaluates all possibilities of measurement error that may influence the true score. In addition, it includes interactions between persons and other factors potentially affecting the measurement (Medvedev et al., 2021; Truong et al., 2020). Any distinct sources of variance in test scores are referred to as facets; facets may be the persons tested (P), the test items (I), and the testing occasion (O), as well as their interactions (Medvedev et al., 2021). In G-theory, the variance related to persons tested is considered the true variance, and it is referred to as the differentiation facet with other facets, such as items and occasions considered sources of measurement error (Paterson et al., 2018). Further explanation of facets is discussed in the data analyses subsection of chapter 4.

G-theory utilizes analysis of variance (ANOVA) for estimating error variances caused by each measurement facet. In G-theory, a single measurement of behaviour (such as item score, subscale score, rating) is theorized as a sample

from a universe of acceptable observations for targeted objects of measurement, denoted as persons. Attributes of the assessment such as individual items of a scale, block of items, prompts, raters or occasions are represented as facets (Vispoel et al., 2018). It is similar to the factor in an ANOVA model.

The Generalizability coefficient (G-coefficient) in G-theory is similar to the reliability coefficients in CTT; the main difference between these two is that G-coefficient is estimated by considering individual sources of measurement error (Vispoel et al., 2018). It can be explained using the following equation:

$$\text{G-coefficient} = \frac{Universe{-}score\ Variance}{Universe{-}score\ Variance + Individual\ sources\ of\ Error\ Variance} \quad (5)$$

In the Generalizability analysis, individual variance components of each facet such as persons, items or occasions are calculated first by repeated measure factorial analysis of variance (ANOVA) to assess the relative influence of individual error variances on overall measurement error (Medvedev et al., 2017). These sources of error variances are also referred to as "noise". It is then corrected using Whimbey's correction method. Whimbey's correction accounts for the sampling used (e.g., random, fixed or mixed). It does not affect facets drawn from infinite populations such as persons (Paterson et al., 2018). The ratio between variances in scores due to the primary variable being measured and the total observed variance is expressed by a reliability coefficient called an intra-class coefficient (*ICC*), which ranges from 0 to 1. It was initially introduced in CTT, almost similar to the signal-to-noise ratio (*SNR*), which can be extracted from any ANOVA analysis. Equation 6 demonstrates that SNR is equivalent to the square of effect size (ES^2); it denotes the ratio of variance in the variable X (ΔX) to the total variance (σ^2) in the data set (Bloch & Norman, 2012).

$$SNR = ES^2 = \frac{\Delta X^2}{\sigma^2} \quad (6)$$

ICC as per *SNR* definition is defined using the following formula (Medvedev et al., 2017):

$$ICC = \frac{SNR}{1+SNR} \quad (7)$$

In G-theory, *ICC* refers to the G-coefficient, and it is defined as the ratio of observed (true) variance caused by the object of measurement to the total variance of universe scores (observed variance and error variance) (Shavelson et al., 1989).

Vispoel et al. (2018) demonstrated assumptions of G-theory that persons are sampled at random from an intended population of interest, and items and occasions are sampled at random from a broader universe of similar items and occasions, respectively. The G-theory consists of two critical concepts, the object of measurement or facet of differentiation which refers to the true variances between persons, and the facets of generalisation, which implies factors or potential sources of variances affecting the measurement scores (such as occasions and raters). In a single facet design (item or occasion is the only measurement facet of interest), an observed score represents the sum of grand mean and effects of persons, the measurement facet of interest (item or occasion) and the combination of persons and measurement facet. For example, the following equation denotes the repeated measures, random-effects ANOVA model in a person × items (P × I) design with person as the object of measurement and item as the measurement facet of interest:

$$Y_{pi} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (Y_{pi} - \mu_p - \mu_i + \mu)$$

$$Ypi = grand\ mean + person\ effect + item\ effect + person \times task\ interaction\ and\ other\ error \quad (8)$$

where Y_{pi} denotes the score of a person on an item.

The G-coefficient estimates the generalisability of test scores across populations and occasions, and it ranges from 0 to 1. G-coefficient implies the ratio of the observed (true) variance attributable to the object of measurement and the total variance of the universe scores, including observed variance and the error variance (Medvedev et al., 2020). In a single facet design, person × item, G-coefficient is calculated as follows:

$$\text{Generalisability (G) coefficient} = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pi,e}^2}{n_i}} \quad (9)$$

where σ_p^2 is the variances of scores across persons, $\sigma^2_{pi,\,e}$ is the observed score of person × item interaction and related error, and n_i is the number of items.

The index of consistency of scores is based on whether they are used for norm-referenced or criterion-referenced decisions (Vispoel et al., 2018). G-theory differentiates between relative or norm-referenced and absolute or criterion-referenced measurement decisions, whereas CTT only emphasises relative decisions (Webb & Shavelson, 2005). A relative decision aims at the rank ordering of individuals (interpretation of test scores based on ranks or norms), whereas an absolute decision focuses on an individual's performance regardless of others (domain or criterion-referenced interpretations). The relative G-coefficients only explains the variances directly affecting a measurement tool, such as person-occasion and person-item interactions.

On the other hand, the absolute G-coefficient is analogous to the phi (φ) coefficient and accounts for absolute error variance that involves both direct and indirect factors (such as items and occasions) influencing an absolute measure (Medvedev et al., 2020). The main limitation of the reliability index for a single-facet design is that they do not account for and differentiate between the critical sources of measurement errors, which typically impact scores derived from individual variances measures: random-response, specific-factor, and transient (Vispoel et al., 2018). The Random-response error indicates "noise" influencing scores within a specific administration occasion resultant from moment-to-moment variations in mood, attention, memory, and other factors. Specific-factor error reflects the invariable response to items not related to the construct being assessed. Transient error represents stable factors that impact scores within a specific occasion (such as fatigue, illness, and motivation) but are inconsistent across occasions. Vispoel et al. (2018) suggested that reliability is likely to be overestimated if each source of variability is not considered in a measurement.

Therefore, a G-theory two-facet design of person × item × occasion has the advantage of evaluating all possible sources of measurement error. For example, Vispoel et al. (2018) demonstrated that in a two-facet, person × item × occasion design, an observed score with the random-effects ANOVA model could be computed by the following equation:

$$Y_{pio} = \mu + \left(\mu_p - \mu\right) + (\mu_i - \mu) + (\mu_o - \mu) + (\mu_{io} - \mu_i - \mu_o + \mu) + (\mu_{pi} - \mu_p - \mu_r + \mu) + (\mu_{po} - \mu_p - \mu_o + \mu) + (Y_{pio} - \mu_{pi} - \mu_{po} - \mu_{io} + \mu_p + \mu_i + \mu_o - \mu) \quad (10)$$

$$
\begin{aligned}
&Score\ of\ a\ person\ on\ a\ item\ on\ one\ occasion \\
&\quad = mean\ across\ persons, items\ and\ occasions \\
&\quad + person\ effect(p) + item\ effect(i) + occasion\ effect(o) \\
&\quad + item \times occasion\ interaction(i \times o) \\
&\quad + person \times item\ interaction(p \times i) \\
&\quad + person \times occasion\ interaction(p \times o) \\
&\quad + person \times task \\
&\quad \times occasion\ interaction\ and\ other\ error\ (p \times t \\
&\quad \times o, residual\ error)
\end{aligned}
$$

In a two-facet design, the variance of individual scores is given by the sum of variance components of all main effects and their interactions, which can be expressed using the following equation:

$$\sigma^2_{Y_{pio}} = \sigma^2_p + \sigma^2_i + \sigma^2_o + \sigma^2_{pi} + \sigma^2_{po} + \sigma^2_{io} + \sigma^2_{pio,e} \qquad (11)$$

Medvedev et al. (2017) proposed the state component index (*SCI*) to compute the relative ratio of state to trait aspect using the absolute variance due to person × occasion interaction, which accounts for all potential error variances in the data. It is expressed as the following formula:

$$SCI = \frac{\sigma^2_s}{\sigma^2_s + \sigma^2_t} \qquad (12)$$

where $\sigma^2_s = \sigma^2_{po}$ is the noise or error variance caused by person × occasion interaction that impacts trait scores and $\sigma^2_t = \sigma^2_p$ (variance of persons or trait aspect). The error variances associated with person × occasion interaction in a scale score implies a scale's sensitivity to dynamic changes, thereby reflecting a state or dynamic aspect (Medvedev et al., 2017). For example, an *SCI* value of

1.00 indicates no trait aspect and only individual state is measured, which is improbable as the trait is a basic predictor of state (Epstein, 1994). On the other hand, an *SCI* > 0.60 reflects a state measure, and the higher the score reflects the ability to capture dynamic (state) changes. Similar to *SCI*, a trait component index (*TCI*) can be calculated using the following formula:

$$TCI = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_s^2} \quad (13)$$

G-Theory's application consists of two stages: a G (Generalizability) study and a D (Decision) study. The G-study includes estimating the G-coefficient, reflecting the overall reliability and generalizability of the assessment scores obtained across participants and occasions. G-study results enable us to distinguish between scales measuring dynamic (state) and enduring (trait) aspects. The D-study involves experimenting with the measurement design of a scale (e.g., modifying item content) and permits an investigator to evaluate how variations in different facets impact reliability and, thereby, determine the most suitable measurement protocols (Medvedev et al., 2021). This advanced statistical method is increasingly being used to investigate measurement tools' psychometric properties and improve measurements; however, studies utilizing G-theory to distinguish dynamic and enduring symptoms are limited.

State and Trait Distinction using G-theory

The test-retest reliability coefficient is a standard CTT method for evaluating the consistency of assessment scores over time, which is merely a correlation of the total test scores at two points of time. This method uses total scale scores and cannot differentiate between enduring and dynamic symptoms because it does not consider the variability of individual items, effects of occasion, person and their interactions in one omnibus analysis (Medvedev et al.,

2021). For example, if a person depression score improves by 1 point on one symptom and decreases by 1 on another symptom, the total score remains the same, and the true variability of symptoms and reliability of a scale cannot be accurately estimated using the test-retest CTT method.

Generally, structural equation methods such as latent trait-state models can also be used to distinguish between the stable and occasion specific variance as they also use analysis of variance as utilized in G theory (Cole et al., 2005; Prenoveau, 2014). However, unlike G theory, these models were not explicitly designed to estimate reliability and generalisability of assessment scores by considering all sources of error, and their interactions and application of these models have been somewhat limited. To distinguish dynamic and enduring aspects, G-theory utilizes the established procedures (Cardinet et al. 2010; Medvedev et al., 2017) such as two-way repeated-measures ANOVA to estimate the variance due to object of measurement (persons) and sources of error variance due to occasion, item, person-occasion, person-item, and person-occasion-item interactions. The basic assumptions of this analysis are that enduring aspects will be reflected by person variance (e.g. true differences between individuals) and high generalisability of scores across occasions (Arterberry et al. 2014), while dynamic aspects should have low generalizability of scores across occasions and a relatively high amount of variance associated with person-occasion interactions reflecting individual's state (Medvedev et al., 2017).

A study by Medvedev et al. (2017) suggested that a distinction between enduring and dynamic symptoms is necessary to establish true reliability. It is only possible through repeated measures. The test-retest reliability is the universally used method for evaluating consistency in scores of psychological

measures across two occasions; however, it does not explain multiple measurement errors affecting observed scores on a particular testing occasion (Medvedev et al., 2021). It only correlates total scores at two occasions and does not consider variability across individual test items. For instance, if one item score increases by 2 and another decreases by 2, it will not affect test-retest reliability, and dynamic changes reflected by these items that impact reliability remained unnoticed.

G-theory is utilised to explore multiple sources of measurement error and evaluate the generalisability of scores across occasions. The application of G-theory to distinguish state and trait is novel, and only a few studies have utilised G-theory to measure temporal reliability and distinguish enduring and dynamic symptoms in psychological measures. For example, a recent study by Medvedev et al. (2021) applied G-theory to identify dynamic and enduring symptoms of schizophrenia by evaluating the psychometric properties of the widely used Positive and Negative Syndromes Scale (PANSS; (Kay et al., 1987) and found that the scale predominantly measures enduring symptoms. They also distinguished enduring and dynamic symptoms in PANSS in general symptoms, positive symptoms and negative symptoms subscale. The study results reported that all subscales measure enduring symptoms relatively more than dynamic symptoms. The most stable symptoms included in the general subscale were poor attention, mannerisms and posturing; delusions on the positive symptoms subscale; blunted affect; and poor rapport in the negative symptoms subscale.

In addition, Paterson et al. (2017) used G-theory to distinguish between state and trait aspects of depression in children using the widely applied Children Depression Inventory (CDI) (Kovacs, 1992). They found two-thirds of variance

attributed to stable symptoms and showed reasonable sensitivity to occasion-specific symptoms of depression in children. G-theory has also been applied to evaluate psychometric properties of other widely used measures such as the Perceived Stress Scale (PSS; Miller et al. (2020) and the Rivermead Post Concussion Symptoms Questionnaire (RPQ; Medvedev et al. (2018). However, no studies have used G-theory to evaluate temporal reliability and distinguish between dynamic and enduring symptoms of depression using the GDS-15 scale.

Aim of Present Study

An accurate distinction of dynamic and enduring components in a measurement is necessary to diagnose and monitor depression as well as to develop appropriate treatment interventions. The most common approach to evaluating reliability is CTT, centred at test-retest reliability coefficients. While the CTT method supposes all possible sources of measurement error as a single factor, the G-theory assesses all potential sources of measurement error, including scale items, assessment occasions and their interactions. Therefore, G-theory is also a suitable method for distinguishing between dynamic and enduring symptoms and estimating the overall reliability of test scores and specific sources of measurement error. The GDS-15 is a widely used assessment tool to measure depression symptoms in the older population, and it has appropriate psychometric properties. However, its temporal reliability of scores and the ability to distinguish between dynamic and enduring symptoms of depression have not been examined yet using the appropriate methodology. Therefore, the present study aimed to apply G-theory in a novel way at the item and scale level to partial out potential sources of variance and distinguish enduring and dynamic symptoms of depression assessed by the GDS-15 with a non-clinical sample of older adults.

Chapter 4 G-study Methods and Results

Purpose

The present study aims to investigate the temporal reliability of GDS-15 utilising G-theory and distinguish dynamic and enduring symptoms assessed by the GDS-15 items. The study also purports to identify all possible sources of measurement error affecting the reliability and generalizability of scores. The study utilised longitudinal data from older adults collected over ten years. The G-theory analysis involved a G-study and D-study. G-study examined the sources of measurement errors or error variances and the overall generalizability of the GDS-15. The D-study investigated the psychometric properties of all items in the GDS-15 and involved experimenting with the measurement design. The Human Research Ethics Committee of the University of Waikato approved the study, ethics approval application number HREC(Health)2020#41 (see Appendix B).

Participants

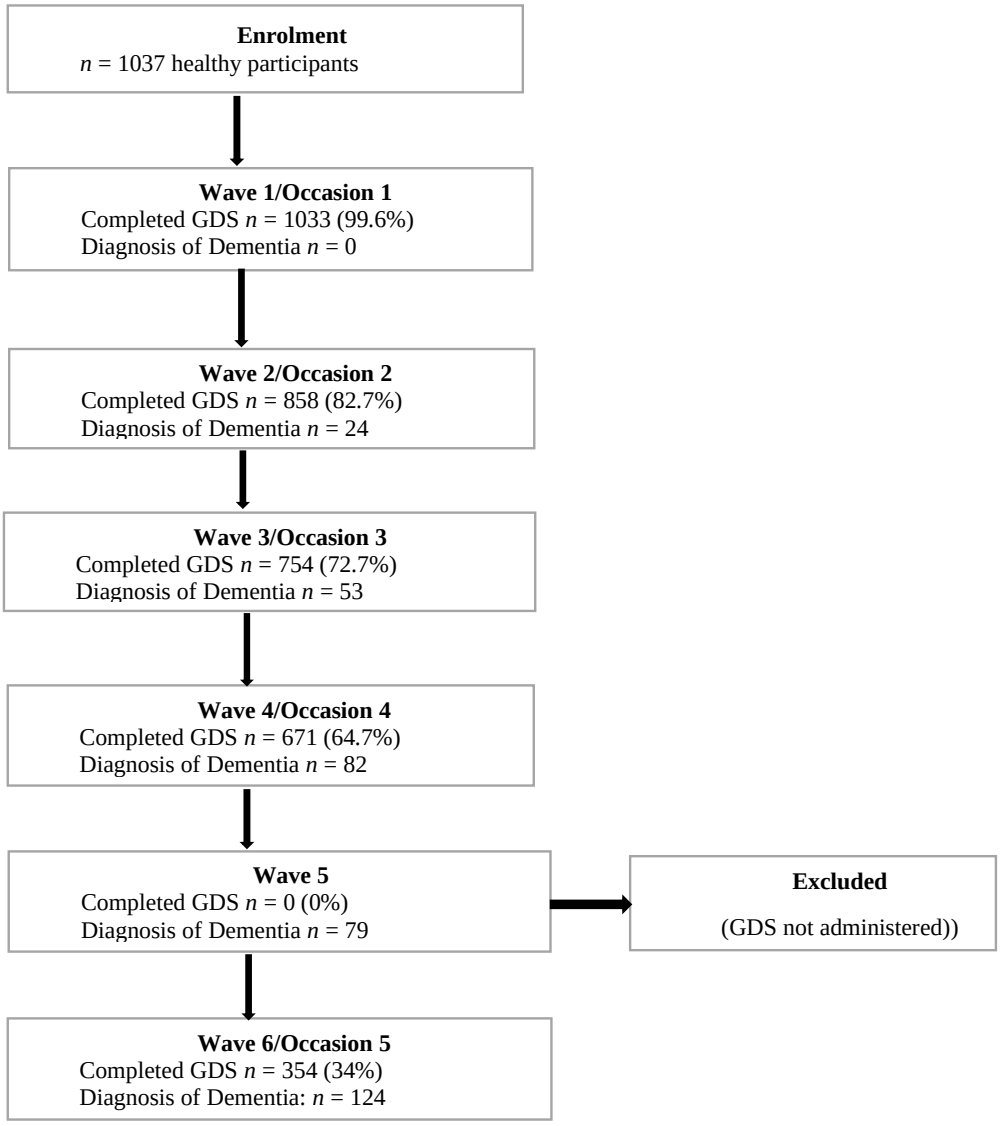
At baseline, participants were older adults aged 70 to 90 years residing in the community in the Eastern suburban areas of Sydney, Australia, who were invited to participate in the Sydney Memory and Ageing Study (MAS) via the electoral roll (Sachdev et al., 2010). The MAS study examined the prevalence of mild cognitive impairment (MCI) and psychological and behavioural symptoms in the non-demented older population (Sachdev et al., 2010). Studies suggest that GDS-15 is unsuitable for people with cognitive and communication difficulties (Azulai & Walsh, 2015), and the reliability of GDS-15 test scores varies with impaired cognitive functioning (Quiroz et al., 2020). However, the validation studies of GDS-15 demonstrated it to be a valid tool in non-demented populations and people with mild cognitive impairment making it an appropriate tool for the MAS study (Nyunt et al., 2009; Quiroz et al., 2020).

Inclusion criteria at baseline were English proficiency sufficient to complete psychometric assessments and self-report questionnaires. Exclusion criteria of the MAS study were any major psychiatric illnesses, acute psychotic symptoms, recent diagnosis of multiple sclerosis, motor neuron disease, developmental disability, progressive malignancy, and a current or previous diagnosis of dementia. Out of 8,914 adults invited, 1,037 participants were eligible and recruited to form the baseline sample (Occasion 1, Wave 1). Participants were assessed biennially, with each assessment being referred to as a "Wave" and each time the GDS was administered being referred to as an "Occasion". Because the GDS was not administered at Wave 5 but was administered at Wave 6, the Wave 6 administration is referred to as Occasion 5 of the GDS. Written consent was sought from all participants and informants involved in this study, and the study was approved by the University of New South Wales Human Ethics Review Committee (HC 05037, 09382, 14327). A more detailed description of recruitment methods and baseline demographic data has been described earlier by Sachdev et al. (2010). Figure 1 represents the consort diagram that includes the number of participants who completed the GDS-15 on each occasion and the total number of participants diagnosed with dementia at each wave.

Of the 1037 participants, 354 (35%) had completed the GDS-15 on all five occasions across the six Waves (10 years) and thus were included in the present G-analyses. Participants who did not complete the GDS-15 on at least two occasions (n = 655) were excluded from the analyses. Data could be missing at each wave due to the following reasons: either the participant was not contactable or could not complete the assessment at that wave, the participant had withdrawn from the study or died, or the participant had advanced dementia and could not

answer questions in later Waves. In the data included for analyses, less than 1% of responses to individual items were missing, so they were imputed using mean imputation for the respective Wave (Huisman, 2000). The sample size of 354 used for the current study surpassed the required sample size for repeated measure ANOVA over five occasions to achieve the power (1-β) of 0.95 to detect a small effect size of 0.15 at a p-value ≤ 0.05.

Figure 1 *CONSORT diagram showing n (%) of participants who completed GDS at each Wave/Occasion and n of participants with a diagnosis of dementia at each Wave/Occasion.*



Measure

The Geriatric Depression Scale: GDS-15

The GDS-15 (Sheikh & Yesavage, 1986) was designed to measure depression in the ageing population. The scale contains 15 dichotomous questions with

response options "yes" or "no" about how a participant has felt over the past week (e.g., "Are you basically satisfied with your life?"), with "yes" scored as 1 and "no" scored as 0. Scores are summed across the 15 items and range from 0 - 15, with a score of 0 – 4 suggesting the absence of depressive symptoms, a score of 5 – 8 indicating mild depressive symptoms, a score of 9 to 11 indicating moderate depressive symptoms and a score of 12 or more indicating severe depressive symptoms. (Sheikh & Yesavage, 1986). In the current sample, Cronbach's alpha was 0.79, which is consistent with previous reports (Gana et al., 2016; Marwijk et al., 1995).

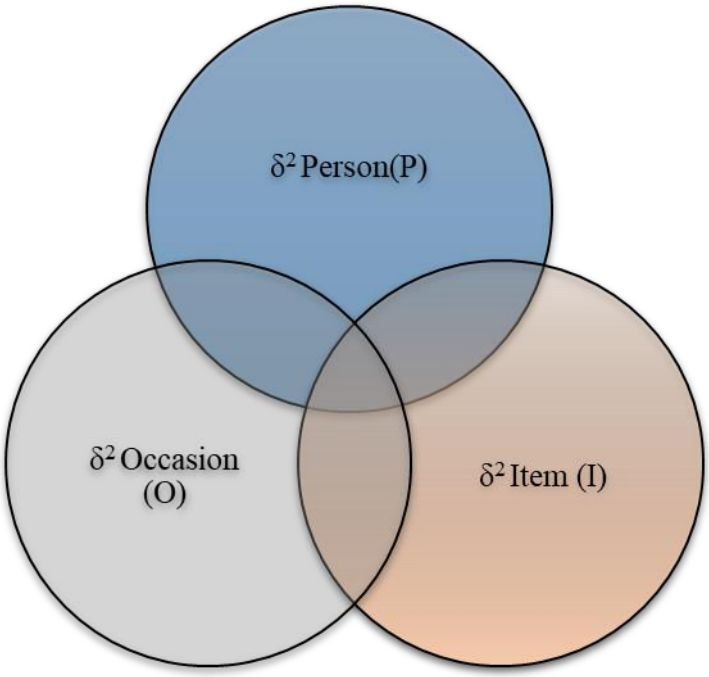
Participants completed the GDS as part of a questionnaire packet mailed to them prior to the face-to-face assessment. If required, incomplete or missing data were collected on the assessment day by the research assistant. In the MAS study, Item 9 of the GDS added the phrase "at night", consistent with the Brink (1982) methodology, such that the question read, "Do you prefer to stay at home *at night*, rather than going out and doing new things?", where the original version of the question read "Do you prefer to stay at home rather than going out and doing new things?". To determine whether this would affect the classification of depression, statistical analysis was performed on two data sets from the same sample—the GDS-15 consistent with the version used in MAS and the GDS-15 consistent with the more common version, which omits 'at night'. Applying the identical cut-point for depression (score ≥6), the results showed a high level of agreement among the two versions (kappa = .931) with a total of only four false positives and no false negatives out of a sample of 354 in the MAS study for validating this item (Sachdev et al., 2010). Appendix C provides the description and scoring of GDS-15.

Data Analyses

Descriptive statistics, including means, standard deviation (*SD*), repeated-measures ANOVA over time, and reliability coefficients, including Cronbach's alpha and the intraclass correlation coefficient (*ICC*), were calculated using IBM SPSS 27 software. The G-theory analyses were conducted using EduG 6.1-e software (Swiss Society for Research in Education Working Group, 2006), following the guidelines explained in Cardinet et al. (2010) and Truong et al. (2020). Both G-study and D-study involved a two-facet design (person × item × occasion). Facets item (I) and occasion (O) were the two facets of interest, also called instrumentation facets, and person (P) was the object of measurement, also called the differentiation facet; these were expressed as P × I × O (Cardinet et al., 2010; Truong et al., 2020; Vispoel et al., 2018). Facet I was fixed given that the same items of assessments were used among all participants involved in the study and across all occasions, whereas the P and O facets were defined as infinite.

Importantly, facet P was not a source of measurement error but the object of measurement. Facet person (P) is the person universe score, item (I) is the item effect, and occasion (O) is the occasion effect. P × I represents the interaction between person (P) and item (I), averaged over occasions. P × O is the interaction effect of the person (P) and occasion (O), averaged over items (I), and indicates the individual state. I × O represents the effect of interaction among items and occasions. P × I × O, e is the interaction between person, item and occasion, including a random error e. Figure 2 graphically illustrates variance components of person, item and occasion and their respective interactions as overlapping areas.

Figure 2 *Venn diagram representing variance components (δ^2) of the person (P) x item (I) x occasion (O) observational design including interactions.*



In a study applying the G-theory method, all sources of error variances are regarded as 100% after controlling for person variance (P), which indicates true differences between persons (Cardinet et al., 2010). G-Theory estimates for the design person by item by occasion, also stated as P × I × O were computed using the formulas in Supplementary Table S1 (Shavelson et al., 1989). The G-study estimates a G-coefficient, which involves a factorial ANOVA corrected for the type of sampling included, and it reflects the generalizability of test scores across persons and occasions (Medvedev et al., 2020). Essentially, G-coefficient is a ratio of true (person) variance to the total variance in the data, including measurement errors (Cardinet et al., 2010). It was demonstrated empirically that ANOVA can be applied for measurement designs using binary ordinal variables with response categories such as 0 (e.g., absence of a symptom) and 1 (e.g., presence of a symptom), which requires at least 40 degrees of freedom

irrespective of distribution and 20 degrees of freedom if at least 20% of respondents endorsed either category (Luney, 1970). These conditions are satisfied with the current sample data.

There are two reliability coefficients for the object of measurement (person) in a generalizability study, relative G-coefficient (G_r) and absolute G-coefficient (G_a). The relative measurement model is based on a norm-referenced approach in which a person's assessment score is compared against others' scores (Vispoel et al., 2018). G_r accounts for relative error variance (δ^2), which directly impact on the P facet (object of measurement) such as person and occasion interaction P × O, and person and item interaction P × I, adjusted for facet sample sizes (Shavelson et al., 1989). G_a is similar to the commonly used Phi (φ) coefficient and accounts for absolute error variance that includes both direct and indirect factors such as items and occasions interaction influencing an absolute measure (Medvedev et al., 2020). Both G coefficients estimate the reliability of a stable pattern of measurement when the person (P) is a differentiation facet. G_r of 0.80 or higher reflects good reliability and generalizability of assessment scores (Cardinet et al., 2010), whereas G_a above 0.70 is determined as acceptable reliability (Arterberry et al., 2014; Truong et al., 2020).

Both state component index (*SCI*) and trait component index (*TCI*) was computed to represent the variance proportion attributed to a dynamic and an enduring pattern in a measure (Medvedev et al., 2017a). *SCI* of 0.60 or higher (*TCI*<0.40) indicates that variance reflects dynamic patterns. On the other hand, *TCI* above 0.60 (*SCI*<0.40) implies that a variance reflects enduring patterns. D-study was utilised to modify the initial measurement tool used for the G study. In the D-study, variance components were calculated for each separate item, and the

effects of removing facets levels were examined to enhance the reliability of the GDS-15.

Results

Participants

The demographic characteristics of the total sample and sub-sample of the G-study are presented in Table 1, which does not include ethnicities due to homogeneity of both samples, including 98% of Caucasian, 1% Asians and 1% Indigenous Australians. In addition, the age range and proportion of males and females and education levels were comparable between the full sample and the subsample used for the current study. All 354 participants included in this study completed the GDS-15 on all five occasions across the six Waves (10 years). Participants who did not have GDS data on all five occasions were excluded from analyses due to the requirement of complete data for the G-theory application. At baseline, the current sample (n=354) did not differ significantly in terms of demographics compared to the full or excluded sample. Therefore, the sample size of 354 used for the current study surpassed the required sample size for repeated measure ANOVA over five occasions to achieve the power (1-β) of 0.95 to detect a small effect size of 0.15 at a p-value ≤ 0.05.

Table 2 *Demographic characteristics of original sample and sub-sample of G-study*

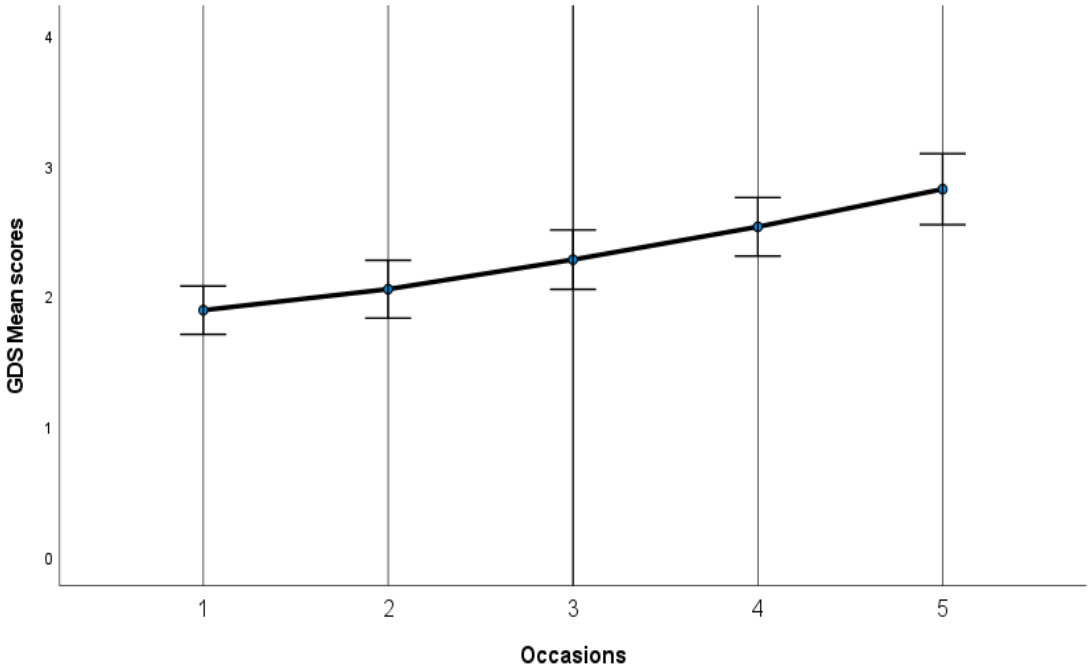
Demographic characteristics	**Original sample (*N*)**	**G-study sample (*n*)**
	1037	354
Age: Mean (SD)	78.84 (4.84)	76.71 (4.12)
Sex		
Male	465	137
Female	572	217
Education		
Primary School	26	9
Incomplete High School	411	131
Completed High School	142	45
Incomplete Tertiary	35	17
Completed Tertiary	311	112
Incomplete High School + Certificate/Diploma	57	21
Complete High School + Certificate/Diploma	55	19
Depression GDS scores: Mean (SD)	2.27 (2.06)	1.86 (1.75)

GDS-15

Figure 3 shows GDS-15 mean scores and 95% confidence intervals (CI) for all assessment Occasions. Repeated measures ANOVA revealed there was a significant main effect of assessment occasion (F (353,1) = 23.98, $p < 0.001$), indicating a significant increase in depression scores in the current sample over time (*ICC* = 0.56 (95% CI [0.51, 0.60], over 5 Occasions). Repeated ANOVA also revealed there was a statistically significant increase in the GDS-15 from Occasion 1 (baseline) to all other Occasions ($p < 0.001$) except Occasion 2 (Bonferroni corrected). In the current sample, McDonald's omega was 0.72 while Cronbach's alpha coefficients across Occasions ranged from 0.73 to 0.79 indicating good internal consistency of the GDS-15 and are consistent with previous reports (Gana et al., 2016; Marwijk et al., 1995). Although the mean depression scores were below 4 in the current sample, indicating participants did not meet the clinical threshold for moderate or severe depression, individual

scores ranged from 0 to 14, and the percentage of participants with depressive symptoms (scores > 4) varied from 8% to 12% across waves. Therefore, this sample included a range of persons with whom the instrument would be useful.

Figure 3 *GDS-15 mean scores and 95% Confidence Intervals (CI) across 5 occasions (n = 354).*



G-Study

The results of G analyses are presented in Table 3. In the G-study, the person (P) is a differentiation facet treated as independent. The interaction between person and occasion (P × O) indicates an individual state or dynamic changes at individual levels. In contrast, occasion (O) indicates the overall effect of the occasion on all individuals (e.g., temporal increase or decrease of depression scores). The GDS-15 scale demonstrated strong reliability of test scores and generalizability of scores across occasions and sample populations (G_a=0.82, G_r=0.90) in measuring enduring depression symptoms. This means that true differences in depression levels across individuals accounted for 82% of the

total variance, with merely 18% of the variance attributed to measurement error. Just over half of the error variance (9.3%) was explained by dynamic symptoms of depression and another half (8.7%) by occasion, which reflected a significant increase in depression scores over time. The EduG analyses outputs for the total GDS-15, including observation and estimation design, ANOVA and G-study table, are presented in Appendix D1. Appendix D1 represents error variance out of 100% by controlling for true person variance while the study tables present all variance components including person true variance out of 100% for convenience of a reader.

Table 3 *G-study estimates of the GDS-15 including variance components with standard errors (SE), G absolute (Ga) and G relative (Gr) coefficients for the Person (P) x Occasion (O) x Item (I) design.*

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.012		.....		.....	82.0
	.....	I	.....		(0.000)	0.0
	.....	O	.....		0.001	8.7
	.....	PI	(0.000)	0.0	(0.000)	0.0
	.....	PO	0.001	10.0	0.001	9.3
	.....	IO	.....		(0.000)	0.0
	.....	PIO	(0.000)	0.0	(0.000)	0.0
Sum of variances	0.012		0.001	100%	0.003	100%
Standard deviation	0.112		Relative SE: 0.038		Absolute SE: 0.053	
Gr	0.90 95% CI [0.83, 0.97]					
Ga	0.82 95% CI [0.72, 0.92]					

Note: Grand mean for levels used: 0.154. Absolute error variance estimated as 100% and computed after accounting for the true person variance, which is the object of measurement. Numbers presented in parentheses signify variance components that are statistically negligible.

D-Study

D-study was designed to identify items that were sensitive to dynamic changes. Table 4 represents variances by person (enduring or trait symptoms), P × O interactions (dynamic or state symptoms), and *SCI* values to determine which GDS-15 items reflect dynamic symptoms. The *SCI* values ranged from 0.32 to 0.70, with a mean of 0.50, suggesting that enduring symptoms of depression mainly explain the variance. Of the 15 items, only three items, "Have you dropped many of your activities and interests" (Item 2), "Do you often feel helpless" (Item 8) and "Do you feel that your situation is hopeless" (Item 14) presented a high *SCI* (≥ 0.60), reflecting sensitivity to dynamic changes. On the contrary, the other 12 items had *SCI*s below 0.60, indicating more enduring patterns of symptoms within the benchmark.

Additional analyses were conducted by removing items with the highest *SCI* values to check if this would further increase *G* coefficients. The items with the highest *SCI* were removed, one by one, and the *G* coefficients were analysed; however, no noticeable increase was observed (Table 4). The current study also examined if the most dynamic symptoms could be measured reliably if combined into one subscale, including Items 2, 8, and 14. However, the resultant measure failed to satisfy the relevant criteria of $SCI \geq 0.60$. Removing each occasion one at a time did not result in a noticeable change of *G* coefficients, suggesting that none of the occasions affected our results and supported the measurement's robustness. All the EduG analyses output of D-study for each of the GDS-15 items, including observation and estimation designs, are included in Appendix D2-D16.

Table 4 *Variance components of Person (P), Occasion (O) and P x O interaction, along with State Component Index (SCI) and G-coefficients for each item contents in the GDS-15 scale (Sheikh & Yesavage, 1986).*

No	Item content	P	O	PxO	*SCI*	*Gr/a*
1	Life satisfaction	0.021	0.001	0.012	0.36	0.63/0.61
2	Interest in daily activities	0.009	0.012	0.021	0.70	0.31/0.22
3	Feeling that life is empty	0.014	0.01	0.017	0.55	0.46/0.35
4	Boredom	0.015	0.001	0.012	0.44	0.54/0.52
5	Good spirit	0.017	0.012	0.019	0.53	0.47/0.36
6	Expectation of negative events	0.017	0.011	0.015	0.47	0.54/0.39
7	Happiness	0.015	0.001	0.013	0.46	0.52/0.50
8	Helplessness	0.014	0.012	0.021	0.60	0.39/0.29
9	Reluctance to go out at night	0.017	0.011	0.016	0.48	0.53/0.39
10	Perceived memory problems	0.02	0.001	0.012	0.38	0.63/0.61
11	Appreciation to be alive	0.014	0.012	0.02	0.59	0.41/0.30
12	Worthlessness	0.018	0.01	0.016	0.47	0.52/0.41
13	Energy	0.025	0.001	0.012	0.32	0.67/0.65
14	Hopelessness	0.011	0.012	0.021	0.66	0.35/0.26
15	Perceived inferiority	0.019	0.009	0.016	0.46	0.55/0.44
	Removing Item 2	0.013	0.001	0.001	0.07	0.90/0.82
	Removing Items 2 and 14	0.013	0.001	0.001	0.07	0.89/0.82
	Removing Items 2, 14 and 8	0.013	0.001	0.001	0.07	0.88/0.81
	Combining Items 2, 14 and 8	0.007	0.012	0.004	0.36	0.42/0.25
	Removing occasion 1	0.012	0.002	0.002	0.14	0.86/0.77
	Removing occasion 2	0.011	0.002	0.002	0.15	0.85/0.73
	Removing occasion 3	0.011	0.002	0.002	0.15	0.87/0.77
	Removing occasion 4	0.014	0.001	0.002	0.13	0.90/0.83
	Removing occasion 5	0.014	0.002	0.002	0.13	0.88/0.79

Chapter 5 Discussion

Main Findings of G-study

This study aimed to investigate the temporal reliability of the GDS-15 and its ability to distinguish between dynamic and enduring symptoms of depression uniquely using G-theory analyses. To date, no other studies have used G-theory to investigate depression in older adults measured by the GDS-15. Using a longitudinal dataset of an adequate size collected over ten years, this study demonstrated that the GDS-15 has strong reliability of test scores and generalizability for assessment scores (G_a = 0.82, G_r = 0.90) across sample population and occasions and measures predominantly enduring symptoms of depression, which makes it a good tool for monitoring impacts of clinical interventions. In addition, there were no other observable sources of measurement error that may have impacted GDS-15 scores. Therefore, the results showed that the GDS-15 is a reliable measure for enduring depression symptoms in the current measurement design.

The present study's findings are consistent with several other large-scale studies of depressive symptoms in older persons (Davey et al., 2004; Gana et al., 2016) that are specific to the nature of their sample. Some of the first papers on G-theory used binary outcome measures as item responses (1, 0) to demonstrate how G-theory worked and demonstrated no impact of using such variables on the accuracy of the results (Brennan, 2001). Binary variables should not be understood as categorical in this case because 1 corresponds to the presence of depressive symptoms and 0 corresponds to its absence for each GDS-15 item. As a parametric approach, G-theory assumes normality of distribution (Cardinet et al., 2010), meaning that it would be less appropriate for the clinically depressed sample. The current sample was optimal for establishing reliability and

generalisability of GDS-15 assessment scores across occasions and sample populations.

Prior studies on the reliability of the GDS-15 utilised test-retest reliability coefficients to evaluate its temporal reliability and reported correlation between two tests at about 0.60, which suggests that GDS-15 predominantly measures dynamic symptoms of depression. However, this score is misleading because it does not account for variability of individual items, impacts of occasion and their interactions. Therefore, the current study used G-theory to examine what aspects of depression were measured by GDS-15 over multiple occasions by accounting for different sources of measurement error, which permitted to establish true reliability and generalisability of GDS scores.

The results of this study have shown that the GDS-15 predominantly measures enduring depression symptoms, which is in line with those reported by Gana et al. (2016), who examined the nature of depression (state vs trait) using the GDS-15 – but using the latent trait-state modelling – in a sample of 753 older adults in France across four occasions. They found that depression assessed by the GDS-15 reflected enduring symptoms of depression in their sample. The same methodology was applied to the Center for Epidemiologic Studies Depression scale (CES-D) completed by older adults across four occasions and demonstrated that enduring aspects (trait factor) explained two-thirds of the total variance (Davey et al., 2004). However, latent trait-state modelling is not suitable to examine the overall reliability and generalizability of assessment scores and specific sources of measurement error, which is a unique feature of the G-theory (Cardinet et al., 2010; Shavelson et al., 1989).

By applying G-theory, the results aligned with the findings above and revealed that 82% of the variance in the current sample was attributable to enduring individual (trait) differences. In comparison, the remaining 18% of the variance was equally split between dynamic changes in depression and the increase of depression scores over time, both of which were statistically significant. Unlike structural equation modelling methods, which do not account for different sources of error variances associated with state-trait variability, this G-study on GDS-15 examined various sources of measurement error such as the items, persons, occasions, and their interactions, demonstrating its applicability and usefulness in research and practice.

Although, the GDS-15 predominantly measures enduring depression symptoms explaining 82% of variance in depression scores, 9.3% of variance was attributed to dynamic (state) aspect of depression. This needs to be considered when monitoring depression levels over time and/or using the GDS-15 for diagnostic purposes. In addition, assessment occasion that also represents a time factor in this study accounted for 8.7% of variance and reflects the group effect of occasion that may contain an invariance error. Considering occasion effect is important if measuring the outcomes because it represents the overall effect on all participants, which may differ across samples and depend on assessment time and place.

The number of studies that use G-theory to investigate enduring and dynamic symptoms of depression is limited to date, which does not allow for comparison between studies. To date, only Paterson et al. (2017) applied G-theory to distinguish between dynamic and enduring symptoms of depression in children aged 9 to 14 using the CDI-10 across three occasions and found acceptable

generalizability (G = 0.79) of this measure, which is similar to the results of present study. In contrast to these results, however, they found that the CDI-10 also measures dynamic aspects of depression in children, with 33% of variance attributed to dynamic (state) aspect and only 66% enduring (trait) aspect of depression) (Paterson et al., 2017). A possible explanation for the result may be that significant dynamic variance (about one-third) could be attributed to rapid developmental changes in children of this age.

Similarly, a recent study used G-theory to examine dynamic and enduring symptoms in a clinical sample of adults with schizophrenia, assessed over five occasions using the PANSS (Kay et al., 1987), found that this measure captured enduring symptoms to a greater extent compared to dynamic symptoms (Medvedev et al., 2020). They also found that depression emerged as an enduring and stable trait variable. These findings are consistent with multiple clinical observations showing depression is stable across different ages and co-morbidities and depends more on its biological rather than volatile external factors. In addition, present findings are consistent with other studies showing persistent cognitive deficits, neuropsychological deficits in executive functioning, and other stable factors such as personality and impaired ability to perform daily activities impact depressive symptoms (Baune et al., 2018; Conradi et al., 2011). These findings provide evidence for the benefits of applying G-theory for understanding depressive symptoms, and more studies on various depression scales in adult populations are necessary to better understand dynamic and enduring depression symptoms.

The results indicated a significant increase in the GDS-15 depression scores over time, consistent with other studies related to ageing, functional

decline, and isolation, instead of the clinical construct of depression (Meeks et al., 2011; Rinaldi et al., 2003). For example, studies have found mild depression or dysthymia in older people two or three times more prevalent than major depression (Meeks et al., 2011; Rinaldi et al., 2003). Moreover, almost 10% of older adults with mild depression develop a major depressive episode every year. Studies have found a curvilinear association between age and symptoms of depression where depressive symptoms are at the lowest levels during mid-age and increase later, with the symptoms beginning around sixty years of age (Kessler et al., 1992; Lewinsohn et al., 1991). However, other affective conditions such as anxiety have declined with age (Trollor et al., 2007). Future studies should use G-theory to investigate the ageing population's dynamic and enduring aspects of anxiety and other neuropsychiatric symptoms.

Findings of D-study

By evaluating variance components, the study identified three dynamic symptoms of depression reflected by GDS items measuring the loss of interest in activities (Item 2), hopelessness (Item 14), and helplessness (Item 8). However, these items did not affect the overall strong reliability of the GDS-15 scores, and their removal did not improve the reliability of test scores, suggesting that the scale has an optimal measurement design that does not require modifications. Therefore, the GDS-15 can be used to capture enduring and dynamic depression in older adults, which may assist in targeted interventions and improved clinical diagnoses over time. Additionally, the results suggest that if a psychological intervention can result in significant changes in depression levels, as measured by the GDS-15, such changes are likely to be enduring. This approach will prove helpful in expanding our understanding of the trait and state aspect of depression in the older population.

Unlike CTT methods that cannot detect variability on specific items (that represent depression symptoms), this study utilised G-theory and examined each of the 15 items of the GDS-15 in terms of enduring and dynamic properties. An *SCI* above 0.60 suggests that items in the scale are sensitive to state changes (measuring dynamic components). Only three items were found to reflect dynamic symptoms: "Have you dropped many of your activities and interests?" (Item 2), "Do you often feel helpless?" (Item 8), and "Do you feel that your situation is hopeless?" (Item 14). All items have true intraindividual variance and showed some enduring components. For instance, although 60% of the variance in helplessness reflected dynamic aspects, 40% of variance reflects its enduring aspect. The *SCI* scores suggested that symptoms such as hopelessness, helplessness, and lack of interest in activities are dynamic and changing over time; however, these results need to be confirmed by future research studies. A possible explanation for this could be based on the hopelessness theory of depression, assuming that one may feel helpless and hopeless due to any current adverse life events that potentially lead to depression if such feelings persist for a long time. However, the results of this study suggest that in many cases, helplessness, hopelessness, and lack of interest may be alleviated over time because these feelings are based on current events and are unlikely to be long-standing. Therefore, any approach targeting an individual's symptoms, such as being hopeless, helpless, and lacking interest in activities, potentially increase the effectiveness of a treatment.

All other items representing symptoms such as life satisfaction, boredom, feeling empty, happiness, being afraid of adverse life events, social isolation, being energetic, and self-esteem reflected enduring aspects of depression and could be more related to ageing and increase with age. Similarly, Chiappelli et al.

(2014) found that experiencing negative emotions and distressing thoughts were stable features of depression. These findings are in line with theoretical explanations of depressive symptoms and have significant implications for understanding the theoretical explanations for the aetiology of depression. For instance, these findings are consistent with the cognitive theory of depression, which proposed the idea of stable and enduring schemas leading to depression symptoms such as sadness, lack of motivation, and being afraid of adverse life events and worthlessness. This combination of findings supports the conceptual premise that understanding dynamic and enduring symptoms of depression would be helpful to develop effective interventions and reduce misdiagnosis.

Implications of Study Findings

This study contributed to the existing knowledge about enduring and dynamic symptoms of depression and applications of G-theory for examining the reliability of assessment tools such as GDS-15. These findings will interest researchers applying GDS-15 to assess depression symptoms in older adults. The results of this study have confirmed that GDS-15 is a reliable tool in assessing enduring depression symptoms in the older population, with only a minor proportion of variance attributable to the measurement error. Furthermore, this is the first study applying G-theory to the GDS-15 to distinguish between enduring and dynamic symptoms of depression. The evidence from this study suggests that GDS-15 can be useful for researchers and professionals to monitor risks of depression accurately and determine if an intervention is likely to be long-lasting, as the scale examines enduring aspects of depression.

D-study was conducted to test various assessment designs for optimising the GDS-15 scale. It revealed that the GDS-15 attain the most reliable and

generalizable scores across populations and occasions in the current measurement design without the need for modification. Even though D-study did not reduce the measurement error of the design or improve the measurement, it has extended the knowledge that no better scale can be derived from GDS-15, and it has an optimum measurement.

Another important implication of the current study is identifying three dynamic symptoms of depression, which did not impact on the overall reliability of the scale. These symptoms are more amenable to change and should be the focus of interventions aimed at reducing depression more efficiently. For instance, interventions such as Mindfulness-Based Stress Reduction (MBSR) and Cognitive Behaviour Therapy (CBT) can be utilized to focus on reducing these dynamic depression symptoms on the first place (Hernandez & Overholser, 2021; Hofmann et al., 2010).

MBSR is structured programmes utilising meditation and mindfulness practices to acquire sustained present moment awareness and impede distressing about the future, such as feeling hopeless and helpful and ruminating about the past (Niazi & Niazi, 2011). Moreover, mindfulness will target dynamic symptoms such as helplessness, hopelessness, and interest in activities. If practised properly, it increases positive emotions while focusing on the present moment and moving away from negative experiences of the past and negative expectations for the future (Krägeloh et al., 2019). CBT focusing on decreasing dynamic depressive symptoms such as helplessness, hopelessness, and lack of interest in activities will be more efficient as they are more responsive to early interventions (Hernandez & Overholser, 2021; McGinn, 2000). Furthermore, the present study raises the possibility of reducing issues on the lack of accurate assessment of LLD.

Limitations

The present study is not without limitations. For example, analyses only included 354 participants who completed the scale at all five assessment occasions, approximately 30% of the total sample at baseline (n = 1037), which may have led to selection bias. The sample was also unrepresentative of the general older adult population as most of the sample consisted of White Europeans (around 98%), and only 39% of the sample were male. This sample is also affluent because the suburbs where the participants were recruited had relatively high education levels, and women worked more than other similar-aged cohorts. Therefore, our findings may have limited generalizability to older adults of other ethnic groups and cultures (Gana et al., 2016), and further studies with more diverse samples are warranted.

Another limitation of the present study is the possibility that the results may be biased because participants with more severe cognitive impairment might have been unable to complete the GDS or limited by lack of insight as the disease progressed. Consistent with previous findings on the prevalence of depression in the community-dwelling older adult populations, the mean depressions scores were low (i.e., < 4) and did not meet the criteria for high/severe levels of depression. As a result, the sample and analyses may not represent the general older Australian population that includes clinical population. Further studies that include participants with clinically significant depressive symptoms are required to explore enduring and dynamic symptoms of depression in clinical populations.

A potential limitation of this study is that measurement invariance was not investigated across assessments using alternative statistical methods. However, the G-study results demonstrated that item effect was effectively zero and the

effect of occasion, which implicitly includes fluctuations due potential changes in item loadings or thresholds across assessments, was merely 8.7% of the total variance and had no impact on the overall strong temporal reliability and generalisability of GDS-15 scores.

Directions for Further Research

G-theory can potentially be used to quantify an individual's amount of trait and state depression. However, this will require collecting a large sample of individual assessments over time, similar to a single-subject design to have a sufficiently large sample size of responses to ensure the reliability of scores. For example, weekly assessments over a year will produce a minimum required sample size of about 50 responses per item to ensure the generalizability of results. Nevertheless, this may open a new research avenue to apply G-theory at the individual level. Moreover, further studies that include participants with clinically significant depressive symptoms are required to explore enduring and dynamic symptoms of depression in clinical populations. The current study explored symptoms over a long time with 2-4 years intervals, and future research should focus on applying the same method using shorter intervals between assessments such as weeks or months. This will help investigate what symptoms show more variability across shorter and longer intervals.

Conclusion

Overall, the findings of this study suggest that the GDS-15 is a reliable scale for measuring enduring symptoms of depression and can be used to evaluate the treatment outcomes. The generalizability coefficients of the current G study were above 0.80, reflecting the high generalizability of test scores across the sample population and occasions. In addition, by analysing *SCI* values and

variance components, it has been found that the GDS-15 mainly measures enduring symptoms of depression, and only three items- helplessness, hopelessness, and lack of interest in activities reflected dynamic symptoms. Identification of dynamic symptoms is beneficial as they are responsive to change, and interventions can target reducing symptoms of helplessness, hopelessness, and lack of interest in activities. Therefore, GDS-15 can be utilized as an appropriate tool to measure depression in the older population.

References

Abramson, L. Y., & Alloy, L. B. (2006). Cognitive Vulnerability to Depression: Current Status and Developmental Origins. In *The interpersonal, cognitive, and social nature of depression* (pp. 83-100). Lawrence Erlbaum Associates Publishers.

Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, *96*(2), 358-372. https://doi.org/10.1037/0033-295X.96.2.358

Acosta Quiroz, C. O., García-Flores, R., & Echeverría-Castro, S. B. (2021). The Geriatric Depression Scale (GDS-15): Validation in Mexico and Disorder in the State of Knowledge. *The International Journal of Aging and Human Development*, *93*(3), 854-863.

Agrell, B., & Dehlin, O. (1989). Comparison of six depression rating scales in geriatric stroke patients. *Stroke*, *20*(9), 1190-1194.

Alexopoulos, G. S. (2005). Depression in the elderly. *Lancet*, *365*(9475), 1961-1970. https://doi.org/10.1016/s0140-6736(05)66665-2

Alexopoulos, G. S., Meyers, B. S., Young, R. C., Campbell, S., Silbersweig, D., & Charlson, M. (1997). 'Vascular depression' hypothesis. *Arch Gen Psychiatry*, *54*(10), 915-922. https://doi.org/10.1001/archpsyc.1997.01830220033006

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Arias-Merino, E. D., Ortiz, G. G., Ruvalcaba, N. M. M., Arias-Merino, M. J., Velazquez-Brizuela, I. E., Meda-Lara, R. M., & Morales-Sanchez, A. E. (2012). Depressive symptoms among community-dwelling Mexican

elderly. In E. Abdel-Rahman (Ed.), *Depression in the Elderly* (pp. 27-48). Nova Science Publishers Inc.

Arterberry, B. J., Martens, M. P., Cadigan, J. M., & Rohrer, D. (2014). Application of Generalizability Theory to the Big Five Inventory. *Personality and Individual Differences, 69,* 98–103. https://doi.org/10.1016/j.paid.2014.05.015

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Aziz, R., & Steffens, D. C. (2013). What are the causes of late-life depression? *The Psychiatric clinics of North America*, *36*(4), 497-516. https://doi.org/10.1016/j.psc.2013.08.001

Ban, T. A. (2014*). From melancholia to depression. A history of diagnosis and treatment.* International Network for the History of Neuropsychopharmacology. https://inhn.org/fileadmin/previews_new/From_Melancholia_to_Depression_March_6_2014.pdf

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic, 2nd ed*. Guilford Press.

Barlow, D. H., Durand, V. M., & Hofmann, S. G. (2016). *Abnormal psychology: An integrative approach*. Cengage learning.

Baune, B. T., Malhi, G. S., Morris, G., Outhred, T., Hamilton, A., Das, P., Bassett, D., Berk, M., Boyce, P., Lyndon, B., Mulder, R., Parker, G., & Singh, A. B. (2018). Cognition in depression: Can we THINC-it better? *Journal of Affective Disorders*, *225*, 559-562. https://doi.org/https://doi.org/10.1016/j.jad.2017.08.080

Bebbington, P. (1996). The origins of sex differences in depressive disorder: Bridging the gap. *International Review of Psychiatry*, *8*(4), 295-332.

Bech, P., Allerup, P., Gram, L. F., Kragh-Sorensen, P., Rafaelsen, O. J., Reisby, N., & Vestergaard, P. (1988). The diagnostic Melancholia scale (DMS): dimensions of endogenous and reactive depression with relationship to the Newcastle scales. *Journal of Affective Disorders*, 14(2), 161-170. https://doi.org/10.1016/0165-0327(88)90059-6

Beck, A. T. (1967). *Depression*. New York.

Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment*. University of Pennsylvania Press.

Beck, A. T., Alford, B. A., Beck, M. A. T., & Alford, P. D. B. A. (2014). *Depression*. University of Pennsylvania Press.

Beijers, L., Wardenaar, K. J., van Loo, H. M., & Schoevers, R. A. (2019). Data-driven biological subtypes of depression: systematic review of biological approaches to depression subtyping. *Molecular Psychiatry*, *24*(6), 888–900. https://doi.org/10.1038/s41380-019-0385-5

Benedetti, A., Wu, Y., Levis, B., Wilchesky, M., Boruff, J., Ioannidis, J. P. A., Patten, S. B., Cuijpers, P., Shrier, I., Gilbody, S., Ismail, Z., McMillan, D., Mitchell, N., Ziegelstein, R. C., & Thombs, B. D. (2018). Diagnostic accuracy of the Geriatric Depression Scale-30, Geriatric Depression Scale-15, Geriatric Depression Scale-5 and Geriatric Depression Scale-4 for detecting major depression: Protocol for a systematic review and individual participant data meta-analysis. *BMJ Open*, *8*(12), [e026598]. https://doi.org/10.1136/bmjopen-2018-026598

Benjamin, S., & Steffens, D. C. (2011). Structural neuroimaging of geriatric depression. *Psychiatric Clinics*, *34*(2), 423-435.

Berman, J., & Furst, L. M. (2011). *Depressed older adults: Education and screening.* Springer Publishing Co.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69-77. https://doi.org/https://doi.org/10.1016/S0022-3999(01)00296-3

Blazer D. G. (2003). Depression in late life: review and commentary. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *58*(3), 249–265. https://doi.org/10.1093/gerona/58.3.m249

Blazer, D. G. (2003). Depression in late life: review and commentary. *J Gerontol A Biol Sci Med Sci*, *58*(3), 249-265. https://doi.org/10.1093/gerona/58.3.m249

Blazer, D. G. (2009). Depression in late life: review and commentary. *Focus*, *7*(1), 118-136.

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher*, *34*(11), 960-992. https://doi.org/10.3109/0142159X.2012.703791

Bowlby, E. (2008). *Loss-Sadness and Depression: Attachment and Loss Volume 3* (Vol. 3). Random House.

Brennan, R. L. (2010). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education*, *24*(1), 1-21. https://doi.org/10.1080/08957347.2011.532417

Brodaty, H., Luscombe, G., Parker, G., Wilhelm, K., Hickie, I., Austin, M. P., & Mitchell, P. (2001). Early and late onset depression in old age: different aetiologies, same phenomenology. *J Affect Disord*, *66*(2-3), 225-236. https://doi.org/10.1016/s0165-0327(00)00317-7

Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG.* Routledge/Taylor & Francis Group.

Cardinet, J., Johnson, S., & Pini, G. (2011). Applying generalizability theory using EduG.

Carmelli, D., Swan, G. E., Kelly-Hayes, M., Wolf, P. A., Reed, T., & Miller, B. (2000). Longitudinal changes in the contribution of genetic and environmental influences to symptoms of depression in older male twins. *Psychology and Aging*, *15*(3), 505.

CDC. https://www.cdc.gov/aging/mentalhealth/depression.htm.

Chiappelli, J., Nugent, K. L., Thangavelu, K., Searcy, K., & Hong, L. E. (2014). Assessment of trait and state aspects of depression in schizophrenia. *Schizophrenia Bulletin, 40*(1), 132-142. https://doi.org/10.1093/schbul/sbt069

Chiesi, F., Primi, C., Pigliautile, M., Baroni, M., Ercolani, S., Paolacci, L., Boccardi, V., & Mecocci, P. (2018). Does the 15-item Geriatric Depression Scale function differently in old people with different levels of cognitive functioning? *Journal of Affective Disorders, 227*, 471-476. https://doi.org/https://doi.org/10.1016/j.jad.2017.11.045

Chiesi, F., Primi, C., Pigliautile, M., Ercolani, S., della Staffa, M. C., Longo, A., Boccardi, V., & Mecocci, P. (2017). The local reliability of the 15-item version of the Geriatric Depression Scale: An item response theory (IRT) study. *Journal of Psychosomatic Research, 96*, 84-88. https://doi.org/https://doi.org/10.1016/j.jpsychores.2017.03.013

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and Conceptual Problems With Longitudinal Trait-State Models: Introducing a Trait-State-

Occasion Model. *Psychological Methods, 10*(1), 3–20. https://doi.org/10.1037/1082-989X.10.1.3

Cole, M. G., & Dendukuri, N. (2003). Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *American journal of psychiatry*, *160*(6), 1147-1156.

Collins, P. Y., Patel, V., Joestl, S. S., March, D., Insel, T. R., Daar, A. S., Scientific Advisory Board and the Executive Committee of the Grand Challenges on Global Mental Health, Anderson, W., Dhansay, M. A., Phillips, A., Shurin, S., Walport, M., Ewart, W., Savill, S. J., Bordin, I. A., Costello, E. J., Durkin, M., Fairburn, C., Glass, R. I., Hall, W., … Stein, D. J. (2011). Grand challenges in global mental health. *Nature*, *475*(7354), 27–30. https://doi.org/10.1038/475027a

Conradi, H., Ormel, J., & De Jonge, P. (2011). Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study. *Psychological Medicine*, *41*(6), 1165.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology*, *XVII*(2), 137-163.

Culang-Reinlieb, M. E., Johnert, L. C., Brickman, A. M., Steffens, D. C., Garcon, E., & Sneed, J. R. (2011). MRI-defined vascular depression: a review of the construct. *Int J Geriatr Psychiatry*, *26*(11), 1101-1108. https://doi.org/10.1002/gps.2668

Davey, A., Halverson, C. F., Jr, Zonderman, A. B., & Costa, P. T., Jr (2004). Change in depressive symptoms in the Baltimore longitudinal study of aging. *The Journals of Gerontology. Series B, Psychological Sciences and*

Social Sciences, *59*(6), P270–P277. https://doi.org/10.1093/geronb/59.6.p270

De Gruijter, D. N., & Leo, J. T. (2007). *Statistical test theory for the behavioral sciences*. CRC Press. https://doi.org/10.1201/9781584889595

Disabato, B. M., Morris, C., Hranilovich, J., D'Angelo, G. M., Zhou, G., Wu, N., Doraiswamy, P. M., & Sheline, Y. I. (2014). Comparison of brain structural variables, neuropsychological factors, and treatment outcome in early-onset versus late-onset late-life depression. *Am J Geriatr Psychiatry*, *22*(10), 1039-1046. https://doi.org/10.1016/j.jagp.2013.02.005

Djernes, J. K. (2006). Prevalence and predictors of depression in populations of elderly: a review. *Acta Psychiatrica Scandinavica*, *113*(5), 372-387.

Dow, B., Lin, X., Pachana, N. A., Bryant, C., LoGiudice, D., Goh, A. M., & Haralambous, B. (2018). Reliability, concurrent validity, and cultural adaptation of the geriatric depression scale and the geriatric anxiety inventory for detecting depression and anxiety symptoms among older Chinese immigrants: an Australian study. *International psychogeriatrics*, *30*(5), 735-748.

Dunn, V. K., & Sacco, W. P. (1989). Psychometric evaluation of the Geriatric Depression Scale and the Zung Self-Rating Depression Scale using an elderly community sample. *Psychology and Aging*, *4*(1), 125.

Epstein, S. (1994). Trait Theory as Personality Theory: Can a Part Be as Great as the Whole? *Psychological Inquiry*, *5*(2), 120-122. https://doi.org/10.1207/s15327965pli0502_4

Espinoza, E., & Kaufman, A. H. (2014). Diagnosis and Treatment of Late-Life Depression. *Psychiatric Times*, *31*(10).

Fang, X., Zhang, C., Wu, Z., Peng, D., Xia, W., Xu, J., Wang, C., Cui, L., Huang, J., & Fang, Y. (2018). Prevalence, risk factors and clinical characteristics of suicidal ideation in Chinese patients with depression. *Journal of Affective Disorders*, *235*, 135-141. https://doi.org/10.1016/j.jad.2018.04.027
Fang, X., Zhang, C., Wu, Z., Peng, D., Xia, W., Xu, J., Wang, C., Cui, L., Huang, J., & Fang, Y. (2018). Prevalence, risk factors and clinical characteristics of suicidal ideation in Chinese patients with depression. *Journal of Affective Disorders*, *235*, 135-141. https://doi.org/10.1016/j.jad.2018.04.027
Fiske, A., Wetherell, J. L., & Gatz, M. (2009). Depression in Older Adults. *Annual Review of Clinical Psychology*, *5*(1), 363-389. https://doi.org/10.1146/annurev.clinpsy.032408.153621

Folstein, M., Liu, T., Peter, I., Buell, J., Arsenault, L., Scott, T., & Qiu, W. W. (2007). The homocysteine hypothesis of depression. *Am J Psychiatry*, *164*(6), 861-867. https://doi.org/10.1176/ajp.2007.164.6.861

Forlani, C., Morri, M., Ferrari, B., Dalmonte, E., Menchetti, M., De Ronchi, D., & Atti, A. R. (2014). Prevalence and gender differences in late-life depression: a population-based study. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *22*(4), 370–380. https://doi.org/10.1016/j.jagp.2012.08.015

Forrest, S. J., Siegert, R. J., Krägeloh, C. U., Landon, J., & Medvedev, O. N. (2021). Generalizability theory distinguishes between state and trait anxiety. *Psychological assessment*, *33*(11), 1080-1088.

Friedman, B., Heisel, M. J., & Delavan, R. L. (2005). Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively

intact, community-dwelling elderly primary care patients. *Journal of the American Geriatrics Society, 53*(9), 1570–1576. https://doi.org/10.1111/j.1532-5415.2005.53461.x

Gabrys, J. B., & Peters, K. (1985). Reliability, discriminant and predictive validity of the Zung Self-Rating Depression Scale. *Psychological reports*, *57*(3_suppl), 1091-1096.

Galea, S., Uddin, M., & Koenen, K. (2011). The urban environment and mental disorders: epigenetic links. *Epigenetics*, *6*(4), 400-404. https://doi.org/10.4161/epi.6.4.14944

Gana, K., Bailly, N., Broc, G., Cazauvieilh, C., & Boudouda, N. E. (2017). The Geriatric Depression Scale: does it measure depressive mood, depressive affect, or both?. *International Journal of Geriatric Psychiatry*, *32*(10), 1150–1157. https://doi.org/10.1002/gps.4582

Goldston, D. B., Reboussin, B. A., & Daniel, S. S. (2006). Predictors of suicide attempts: State and trait components. *Journal of Abnormal Psychology, 115*(4), 842–849. https://doi.org/10.1037/0021-843X.115.4.842

Gonçalves-Pereira, M., Prina, A. M., Cardoso, A. M., da Silva, J. A., Prince, M., Xavier, M., & 10/66 Workgroup in Portugal (2019). The prevalence of late-life depression in a Portuguese community sample: A 10/66 Dementia Research Group study. *Journal of Affective Disorders*, *246*, 674–681. https://doi.org/10.1016/j.jad.2018.12.067

Grayson, L., & Thomas, A. (2013). A systematic review comparing clinical features in early age at onset and late age at onset late-life depression. *Journal of Affective Disorders*, *150*(2), 161–170. https://doi.org/10.1016/j.jad.2013.03.021

Guerin, J. M., Copersino, M. L., & Schretlen, D. J. (2018). Clinical utility of the 15-item geriatric depression scale (GDS-15) for use with young and middle-aged adults. *Journal of Affective Disorders*, *241*, 59–62. https://doi.org/10.1016/j.jad.2018.07.038

Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2020). A Meta-Analysis of the Generalization of the Reliability of State/Trait Depression Inventory Scores. *Psicothema*, *32*(4), 476-489. https://doi.org/10.7334/psicothema2020.106

Haigh, E. A. P., Bogucki, O. E., Sigmon, S. T., & Blazer, D. G. (2018). Depression Among Older Adults: A 20-Year Update on Five Common Myths and Misconceptions. *The American Journal of Geriatric Psychiatry*, *26*(1), 107-122. https://doi.org/https://doi.org/10.1016/j.jagp.2017.06.011

Haigh, E., Bogucki, O. E., Sigmon, S. T., & Blazer, D. G. (2018). Depression Among Older Adults: A 20-Year Update on Five Common Myths and Misconceptions. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *26*(1), 107–122. https://doi.org/10.1016/j.jagp.2017.06.011.

Hayward, R. D., Taylor, W. D., Smoski, M. J., Steffens, D. C., & Payne, M. E. (2013). Association of five-factor model personality domains and facets with presence, onset, and treatment outcomes of major depression in older adults. *Am J Geriatr Psychiatry, 21*(1), 88-96. https://doi.org/10.1016/j.jagp.2012.11.012

Hernandez, S. C., & Overholser, J. C. (2021). A Systematic Review of Interventions for Hope/Hopelessness in Older Adults. *Clinical*

Gerontologist, *44*(2), 97-111. https://doi.org/10.1080/07317115.2019.1711281

Hofmann, S. G., Sawyer, A. T., Witt, A. A., & Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 78*(2), 169-183. doi:10.1037/a0018555

Huisman, M. (2000). Imputation of Missing Item Responses: Some Simple Techniques. *Quality and Quantity, 34*(4), 331-351. https://doi.org/10.1023/A:1004782230065

Imai, H., Yamanaka, G., Ishimoto, Y., Kimura, Y., Fukutomi, E., Chen, W. L., ... & Matsubayashi, K. (2014). Factor structures of a Japanese version of the Geriatric Depression Scale and its correlation with the quality of life and functional ability. *Psychiatry research, 215*(2), 460-465.

Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of internal medicine, 159*(15), 1701-1704.

Janssen, J., Hulshoff Pol, H. E., de Leeuw, F.-E., Schnack, H. G., Lampe, I. K., Kok, R. M., Kahn, R. S., & Heeren, T. J. (2007). Hippocampal volume and subcortical white matter lesions in late life depression: comparison of early and late onset depression. *Journal of neurology, neurosurgery, and psychiatry*, *78*(6), 638-640. https://doi.org/10.1136/jnnp.2006.098087

Johnson, W., McGue, M., Gaist, D., Vaupel, J., & Christensen, K. (2002). Frequency and heritability of depression symptomatology in the second half of life: evidence from Danish twins over 45. *Psychological Medicine, 32*(7), 1175-1185.

Jorm, A. F. (2001). History of depression as a risk factor for dementia: an updated review. *Australian & New Zealand Journal of Psychiatry*, *35*(6), 776-781.

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261-276. https://doi.org/10.1093/schbul/13.2.261

Kaya, B. (1999). Late life and depression: Diagnosis and assessment. *Turkish J Geriatr*, *2*(2), 76-82.

Kennedy, G. J. (2015). *Geriatric Depression A Clinical Guide*. The Guilford Press.

Kessler, R. C., Foster, C., Webster, P. S., & House, J. S. (1992). The relationship between age and depressive symptoms in two national surveys. *Psychology and Aging*, *7*(1), 119–126. https://doi.org/10.1037//0882-7974.7.1.119

Koolschijn, P. C. M., van Haren, N. E., Lensvelt‐Mulders, G. J., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: A meta‐analysis of magnetic resonance imaging studies. *Human brain mapping*, *30*(11), 3719-3735.

Kovacs, M. (1992). *Children's depression inventory: Manual* (p. Q8). North Tonawanda, NY: Multi-Health Systems.

Krägeloh, C. U., Henning, M. A., Medvedev, O. N., Feng, X. J., Moir, F., Billington, R., & Siegert, R. J. (2019). *Mindfulness-based intervention research: Characteristics, approaches, and developments*. Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315545875

Kumar, S. S., Merkin, A. G., Numbers, K., Sachdev, P. S., Brodaty, H., Kochan, N. A., Trollor, J. N., Mahon, S., & Medvedev, O. (2022). A novel

approach to investigate depression symptoms in the aging population using generalizability theory. *Psychological Assessment.* Advance online publication. https://doi.org/10.1037/pas0001129

Lambert, K. G., Nelson, R. J., Jovanovic, T., & Cerda, M. (2015). Brains in the city: neurobiological effects of urbanization. *Neurosci Biobehav Rev, 58*, 107–122. https://doi.org/10.1016/j.neubiorev.2015.04.00

Leles da Costa Dias, F., Teixeira, A. L., Cerqueira Guimarães, H., Borges Santos, A. P., Rios Fonseca Ritter, S., Barbosa Machado, J. C., Tonidandel Barbosa, M., & Caramelli, P. (2019). Prevalence of late-life depression and its correlates in a community-dwelling low-educated population aged 75+ years: The Pietà study. *Journal of Affective Disorders*, *242*, 173-179. https://doi.org/10.1016/j.jad.2018.08.012

Leles da Costa Dias, F., Teixeira, A. L., Cerqueira Guimarães, H., Borges Santos, A. P., Rios Fonseca Ritter, S., Barbosa Machado, J. C., Tonidandel Barbosa, M., & Caramelli, P. (2019). Prevalence of late-life depression and its correlates in a community-dwelling low-educated population aged 75+ years: The Pietà study. *Journal of Affective Disorders*, *242*, 173-179. https://doi.org/10.1016/j.jad.2018.08.012

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. A. (1991). Age and depression: Unique and shared effects. *Psychology and Aging*, *6*(2), 247–260. https://doi.org/10.1037/0882-7974.6.2.247

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., Orgeta, V., … Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet

Commission. *Lancet (London, England)*, *396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Lopez, M. N., Quan, N. M., & Carvajal, P. M. (2010). A psychometric study of the Geriatric Depression Scale. *European Journal of Psychological Assessment*.

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Lunney, G. H. (1970). Using Analysis of Variance with a Dichotomous Dependent Variable: An Empirical Study. *Journal of Educational Measurement*, *7*(4), 263–269. http://www.jstor.org/stable/1434469

Luppa, M., Sikorski, C., Luck, T., Ehreke, L., Konnopka, A., Wiese, B., Weyerer, S., König, H. H., & Riedel-Heller, S. G. (2012). Age- and gender-specific prevalence of depression in latest-life – Systematic review and meta-analysis. *Journal of Affective Disorders*, *136*(3), 212-221. https://doi.org/https://doi.org/10.1016/j.jad.2010.11.033

Maier, S. F., & Seligman, M. E. (1976). Learned helplessness: theory and evidence. *Journal of experimental psychology: general*, *105*(1), 3.

Martens, M. P., Parker, J. C., Smarr, K. L., Hewett, J. E., Ge, B., Slaughter, J. R., & Walker, S. E. (2006). Development of a shortened Center for Epidemiological Studies Depression scale for assessment of depression in Rheumatoid Arthritis. *Rehabilitation Psychology*, *51*(2), 135.

McGinn, L. K. (2000). Cognitive behavioral therapy of depression: theory, treatment, and empirical status. *American Journal of Psychotherapy*, *54*(2), 257-262.

Medvedev, O. N., Berk, M., Dean, O. M., Brown, E., Sandham, M. H., Dipnall, J. F., McNamara, R. K., Sumich, A., Krägeloh, C. U., Narayanan, A., &

Siegert, R. J. (2021). A novel way to quantify schizophrenia symptoms in clinical trials. *European Journal of Clinical Investigation*, *51*(3), e13398. https://doi.org/10.1111/eci.13398

Medvedev, O. N., Berk, M., Dean, O. M., Brown, E., Sandham, M. H., Dipnall, J. F., McNamara, R. K., Sumich, A., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2021). A novel way to quantify schizophrenia symptoms in clinical trials. *European Journal of Clinical Investigation*, *51*(3), e13398. https://doi.org/https://doi.org/10.1111/eci.13398

Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017). Measuring mindfulness: Applying generalisability theory to distinguish between state and trait. *Mindfulness*, *8*(4), 1036-1046. https://doi.org/10.1007/s12671-017-0679-0

Medvedev, O. N., Krägeloh, C. U., Narayanan, A., & Siegert, R. J. (2017). Measuring Mindfulness: Applying Generalizability Theory to Distinguish between State and Trait. *Mindfulness*, *8*(4), 1036-1046. https://doi.org/10.1007/s12671-017-0679-0

Medvedev, O. N., Theadom, A., Barker-Collo, S., & Feigin, V. (2018). Distinguishing between enduring and dynamic concussion symptoms: applying Generalisability Theory to the Rivermead Post Concussion Symptoms Questionnaire (RPQ). *PeerJ*. https://doi.org/10.7717/peerj.5676

Meeks, T. W., Vahia, I. V., Lavretsky, H., Kulkarni, G., & Jeste, D. V. (2011). A tune in "a minor" can "b major": a review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *Journal of Affective Disorders*, *129*(1-3), 126-142. https://doi.org/10.1016/j.jad.2010.09.015

Meeks, T., Vahia, I., Lavretsky, H., Kulkarni, G., & Jeste, D. (2011). A tune in 'A minor' can be 'B major': A review of epidemiology, illness course, and public health implications of subthreshold depression in older adults. *Journal of Affective Disorders, 129,* 126-142.

Merkin, A. G., Medvedev, O. N., Sachdev, P. S., Tippett, L., Krishnamurthi, R., Mahon, S., Kasabov, N., Parmar, P., Crawford, J., Doborjeh, Z. G., Doborjeh, M. G., Kang, K., Kochan, N. A., Bahrami, H., Brodaty, H., & Feigin, V. L. (2020). New avenue for the geriatric depression scale: Rasch transformation enhances reliability of assessment. *Journal of Affective Disorders, 264*, 7-14. https://doi.org/https://doi.org/10.1016/j.jad.2019.11.100

Miller, Y. R., Medvedev, O. N., Hwang, Y.S., & Singh, N. N. (2020). Applying generalizability theory to the Perceived Stress Scale to evaluate stable and dynamic aspects of educators' stress. *International Journal of Stress Management*. https://doi.org/10.1037/str0000207

Mizushima, J., Sakurai, H., Mizuno, Y., Shinfuku, M., Tani, H., Yoshida, K., Ozawa, C., Serizawa, A., Kodashiro, N., Koide, S., Minamisawa, A., Mutsumoto, E., Nagai, N., Noda, S., Tachino, G., Takahashi, T., Takeuchi, H., Kikuchi, T., Uchida, H., Watanabe, K., Kocha, H., & Mimura, M. (2013). Melancholic and reactive depression: a reappraisal of old categories. *BMC Psychiatry, 13*(1), 311. https://doi.org/10.1186/1471-244X-13-311

Mohd Sidik, S., Muhd Zulkefli, N. A., & Mustaqim, A. (2003). Prevalence of depression with chronic illness among the elderly in a rural community in Malaysia. *Asia Pacific Family Medicine, 2,* 196-199.

Morse, J., & Lynch, T. R. (2004). A preliminary investigation of self-reported personality disorders in late life: prevalence, predictors of depressive severity, and clinical correlates. *Aging & Mental Health*, *8*(4), 307-315.

Naismith, S. L., Norrie, L. M., Mowszowski, L., & Hickie, I. B. (2012). The neurobiology of depression in later-life: clinical, neuropsychological, neuroimaging and pathophysiological features. *Prog Neurobiol, 98*, 99–143.

Naushad, N., Dunn, L. B., Muñoz, R. F., & Leykin, Y. (2018). Depression increases subjective stigma of chronic pain. *Journal of Affective Disorders*, *229*, 456-462. https://doi.org/10.1016/j.jad.2017.12.085

Ng, T.P. (2018). Old age depression: worse clinical course, brighter treatment prospects? *Lancet Psychiatry, 5* (7), 533–534.

Niazi, A. K., & Niazi, S. K. (2011). Mindfulness-based stress reduction: a non-pharmacological approach for chronic illnesses. *North American journal of medical sciences*, *3*(1), 20–23. https://doi.org/10.4297/najms.2011.320

Niu, L., Jia, C., Ma, Z., Wang, G., Yu, Z., & Zhou, L. (2018). Validating the Geriatric Depression Scale with proxy-based data: A case-control psychological autopsy study in rural China. *Journal of Affective Disorders, 241*, 533–538. https://doi.org/10.1016/j.jad.2018.08.066

Norris, M. P., Arnau, R. C., Bramson, R., & Meagher, M. W. (2004). The efficacy of somatic symptoms in assessing depression in older primary care patients. *Clinical Gerontologist*, *27*(1-2), 43-57.

Nunnally, J. C. (1994). *Psychometric theory 3E*. Tata McGraw-hill education.

Nyunt, M. S. Z., Fones, C., Niti, M., & Ng, T. P. (2009). Criterion-based validity and reliability of the Geriatric Depression Screening Scale (GDS-15) in a

large validation sample of community-living Asian older adults. *Aging and Mental Health*, *13*(3), 376-382.

Paradiso, S., Naridze, R., & Holm‐Brown, E. (2012). Lifetime romantic attachment style and social adaptation in late‐onset depression. *International journal of geriatric psychiatry*, *27*(10), 1008-1016.

Paterson, J., Medvedev, O. N., Sumich, A., Tautolo, E. S., Krägeloh, C. U., Sisk, R., McNamara, R. K., Berk, M., Narayanan, A., & Siegert, R. J. (2018). Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalizability Theory. *J Affect Disord*, *227*, 698-704. https://doi.org/10.1016/j.jad.2017.11.075

Paterson, J., Medvedev, O. N., Sumich, A., Tautolo, E. S., Krägeloh, C. U., Sisk, R., McNamara, R. K., Berk, M., Narayanan, A., & Siegert, R. J. (2018). Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalizability Theory. *J Affect Disord*, *227*, 698-704. https://doi.org/10.1016/j.jad.2017.11.075

Perera, J. K., Rosenblat, J. D., & Flint, A. J. (2017). Dementia with Lewy bodies presenting as psychotic depression. *Aust N Z J Psychiatry*, *51*(11), 1160-1161. https://doi.org/10.1177/0004867417720036

Pocklington, C., Gilbody, S., Manea, L., & McMillan, D. (2016). The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. *Int J Geriatr Psychiatry*, *31*(8), 837-857. https://doi.org/10.1002/gps.4407

Prenoveau, J. M. (2014). Trait-State-Occasion Models. *The Encyclopedia of Clinical Psychology*, 1-9.

Rapp, M. A., Dahlman, K., Sano, M., Grossman, H. T., Haroutunian, V., & Gorman, J. M. (2005). Neuropsychological differences between late-onset

and recurrent geriatric major depression. *Am J Psychiatry, 162*(4), 691-698. https://doi.org/10.1176/appi.ajp.162.4.691

Read, J. R., Sharpe, L., Modini, M., & Dear, B. F. (2017). Multimorbidity and depression: A systematic review and meta-analysis. *J Affect Disord, 221*, 36-46. https://doi.org/10.1016/j.jad.2017.06.009

Reynolds, K., Pietrzak, R. H., El-Gabalawy, R., Mackenzie, C. S., & Sareen, J. (2015). Prevalence of psychiatric disorders in U.S. older adults: findings from a nationally representative survey. *World Psychiatry, 14*(1), 74-81. https://doi.org/10.1002/wps.20193

Rinaldi, P., Mecocci, P., Benedetti, C., Ercolani, S., Bregnocchi, M., Menculini, G., Catani, M., Senin, U., & Cherubini, A. (2003). Validation of the five-item geriatric depression scale in elderly subjects in three different settings. *J Am Geriatr Soc, 51*(5), 694-698. https://doi.org/10.1034/j.1600-0579.2003.00216.x

Sachdev, P. S., Brodaty, H., Reppermund, S., Kochan, N. A., Trollor, J. N., Draper, B., Slavin, M. J., Crawford, J., Kang, K., Broe, G. A., Mather, K. A., Lux, O., & Memory and Ageing Study Team. (2010). The Sydney Memory and Ageing Study (MAS): Methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *International Psychogeriatrics, 22*(8), 1248–1264. https://doi.org/10.1017/S1041610210001067

Sachs-Ericsson, N., Corsentino, E., Moxley, J., Hames, J. L., Rushing, N. C., Sawyer, K., Joiner, T., Selby, E. A., Zarit, S., Gotlib, I. H., & Steffens, D. C. (2013). A longitudinal study of differences in late- and early-onset geriatric depression: depressive symptoms and psychosocial, cognitive,

and neurological functioning. *Aging Ment Health*, *17*(1), 1-11. https://doi.org/10.1080/13607863.2012.717253

Sadock, B. J., Sadock, V. A., & Ruiz, P. (2015). *Kaplan & Sadock's synopsis of psychiatry: Behavioral sciences/clinical psychiatry* (Eleventh ed.). Wolters Kluwer.

Sawyer, R. L. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385-401.

Schulman, E., Gairola, G., Kuder, L., & McCulloch, J. (2002). Depression and associated characteristics among community-based elderly people. *J Allied Health*, *31*(3), 140-146.

Seligman, M. E. P. (1972). Learned Helplessness. *Annual Review of Medicine*, *23*(1), 407-412. https://doi.org/10.1146/annurev.me.23.020172.002203

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalisability theory. *American Psychologist, 44*, 599-612.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, *44*(6), 922.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *5*(1-2), 165–173. https://doi.org/10.1300/J018v05n01_09

Sims, J., O'Connor, D., & Browning, C. (2012). Management of Depression in Older People: A Role for Physical Activity. In E. Abdel-Rahman (Ed.), *Depression in the Elderly* (pp. 115-139). Nova Science Publishers Inc.

Sjöberg, L., Karlsson, B., Atti, A. R., Skoog, I., Fratiglioni, L., & Wang, H. X. (2017). Prevalence of depression: Comparisons of different depression

definitions in population-based samples of older adults. *J Affect Disord*, *221*, 123-131. https://doi.org/10.1016/j.jad.2017.06.011

Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care & Research*, *63*(S11), S454-S466. https://doi.org/https://doi.org/10.1002/acr.20556

Snyder, A. G., Stanley, M. A., Novy, D. M., Averill, P. M., & Beck, J. G. (2000). Measures of depression in older adults with generalized anxiety disorder: A psychometric evaluation. *Depression and Anxiety*, *11*(3), 114-120.

Sözeri-Varma, G. (2012). Depression in the elderly: clinical features and risk factors. *Aging and disease*, *3*(6), 465.

Srivastava, K. (2009). Urbanization and mental health. *Ind Psychiatry J*, *18*(2), 75-76. https://doi.org/10.4103/0972-6748.64028

Steffens, D. C., Skoog, I., Norton, M. C., Hart, A. D., Tschanz, J. T., Plassman, B. L., Wyse, B. W., Welsh-Bohmer, K. A., & Breitner, J. C. (2000). Prevalence of Depression and Its Treatment in an Elderly Population. The Cache County Study. *Archives of General Psychiatry*, *57*, 601-607.

Steunenberg, B., Beekman, A. T., Deeg, D. J., & Kerkhof, A. J. (2010). Personality predicts recurrence of late-life depression. *Journal of Affective Disorders*, *123*(1-3), 164-172.

Taylor, W. D., MacFALL, J. R., Payne, M. E., McQUOID, D. R., Steffens, D. C., Provenzale, J. M., & Krishnan, K. R. R. (2007). Orbitofrontal cortex volume in late life depression: influence of hyperintense lesions and genetic polymorphisms. *Psychological Medicine*, *37*(12), 1763-1773.

Teo, A. R., Marsh, H. E., Forsberg, C. W., Nicolaidis, C., Chen, J. I., Newsom, J., Saha, S., & Dobscha, S. K. (2018). Loneliness is closely associated with depression outcomes and suicidal ideation among military veterans in primary care. *Journal of Affective Disorders, 230,* 42–49. https://doi.org/10.1016/j.jad.2018.01.003

Trajković, G., Starčević, V., Latas, M., Leštarević, M., Ille, T., Bukumirić, Z., & Marinković, J. (2011). Reliability of the Hamilton Rating Scale for Depression: A meta-analysis over a period of 49years. *Psychiatry Research, 189*(1), 1-9. https://doi.org/https://doi.org/10.1016/j.psychres.2010.12.007

Trollor, J. N., Anderson, T. M., Sachdev, P. S., Brodaty, H., & Andrews, G. (2007). Age Shall not Weary Them: Mental Health in the Middle-Aged and the Elderly. *Australian & New Zealand Journal of Psychiatry, 41*(7), 581-589. https://doi.org/10.1080/00048670701392817

Truong, Q. C., Krägeloh, C. U., Siegert, R. J., Landon, J., & Medvedev, O. N. (2020). Applying Generalizability Theory to Differentiate Between Trait and State in the Five Facet Mindfulness Questionnaire (FFMQ). *Mindfulness, 11*(4), 953-963. https://doi.org/10.1007/s12671-020-01324-7

Tsang, R. S. M., Mather, K. A., Sachdev, P. S., & Reppermund, S. (2017). Systematic review and meta-analysis of genetic studies of late-life depression. *Neuroscience & Biobehavioral Reviews, 75,* 129-139. https://doi.org/https://doi.org/10.1016/j.neubiorev.2017.01.028

Tsoi, K. K., Chan, J. Y., Hirai, H. W., & Wong, S. Y. (2017). Comparison of diagnostic performance of Two-Question Screen and 15 depression screening instruments for older adults: systematic review and meta-

analysis. *Br J Psychiatry, 210*(4), 255-260. https://doi.org/10.1192/bjp.bp.116.186932

van Marwijk, H. W., Wallace, P., de Bock, G. H., Hermans, J., Kaptein, A. A., & Mulder, J. D. (1995). Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *The British Journal of General Practice : the Journal of the Royal College of General Practitioners*, *45*(393), 195–199.

Vispoel, W. P., Morris, C. A., & Kilinc, M. (2018). Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychological Methods, 23*(1), 1-26. https://doi.org/10.1037/met0000107

Wang, H. R., Cho, H., & Kim, D. J. (2018). Prevalence and correlates of comorbid depression in a nonclinical online sample with DSM-5 internet gaming disorder. *J Affect Disord, 226*, 1-5. https://doi.org/10.1016/j.jad.2017.08.005

Webb, N. M., & Shavelson, R. J. (2005). Generalizability Theory: Overview. In *Encyclopedia of Statistics in Behavioral Science*. https://doi.org/https://doi.org/10.1002/0470013192.bsa703

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of psychiatric research, 17*(1), 37-49.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

Zung, W. W. K. (1986). Zung Self-Rating Depression Scale and Depression Status Inventory. In N. Sartorius & T. A. Ban (Eds.), *Assessment of*

Depression (pp. 221-231). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-70486-4_21

Zung, W. W., Richards, C. B., & Short, M. J. (1965). Self-rating depression scale in an outpatient clinic: further validation of the SDS. *Archives of general psychiatry*, *13*(6), 508-515.

Appendix A

Published journal article

Appendix B

Ethics Approval

The University of Waikato
Private Bag 3105
Gate 1, Knighton Road
Hamilton, New Zealand

Human Research Ethics Committee
Roger Moltzen
Telephone: +64021658119
Email:humanethics@waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

29 June 2020

Dr Oleg Medvedev
School of Psychology
DALPSS
By email: omedvede@waikato.ac.nz

Dear Oleg

HREC(Health)2020#41: Protective and Risk Factors of Mentally Healthy Aging

Thank you for submitting your amended application HREC(Health)2020#41 for ethical approval.

We are now pleased to provide formal approval for your project.

Please contact the committee by email (humanethics@waikato.ac.nz) if you wish to make changes to your project as it unfolds, quoting your application number with your future correspondence. Any minor changes or additions to the approved research activities can be handled outside the monthly application cycle.

We wish you all the best with your research.

Regards,

Emeritus Professor Roger Moltzen MNZM
Chairperson
University of Waikato Human Research Ethics Committee

Appendix C

Geriatric Depression Scale: Short Form

TARGET POPULATION: The GDS may be used with healthy, medically ill and mild to moderately cognitively impaired older adults. It has been extensively used in community, acute and long-term care settings.

VALIDITY AND RELIABILITY: The GDS was found to have a 92% sensitivity and a 89% specificity when evaluated against diagnostic criteria. The validity and reliability of the tool have been supported through both clinical practice and research. In a validation study comparing the Long and Short Forms of the GDS for self-rating of symptoms of depression, both were successful in differentiating depressed from non-depressed adults with a high correlation ($r = .84$, $p < .001$) (Sheikh & Yesavage, 1986).

Geriatric Depression Scale: Short Form

Choose the best answer for how you have felt over the past week:

1. Are you basically satisfied with your life? YES / NO
2. Have you dropped many of your activities and interests? YES / NO
3. Do you feel that your life is empty? YES / NO
4. Do you often get bored? YES / NO
5. Are you in good spirits most of the time? YES / NO
6. Are you afraid that something bad is going to happen to you? YES / NO
7. Do you feel happy most of the time? YES / NO
8. Do you often feel helpless? YES / NO

9. Do you prefer to stay at home at night, rather than going out and doing new things? YES / NO

10. Do you feel you have more problems with memory than most? YES / NO

11. Do you think it is wonderful to be alive now? YES / NO

12. Do you feel pretty worthless the way you are now? YES / NO

13. Do you feel full of energy? YES / NO

14. Do you feel that your situation is hopeless? YES / NO

15. Do you think that most people are better off than you are? YES / NO

Scoring:

Answers in bold indicate depression. Score 1 point for each bolded answer.

A score > 5 points is suggestive of depression.

A score ≥ 10 points is almost always indicative of depression.

A score > 5 points should warrant a follow-up comprehensive assessment.

References:

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health, 5*(1-2), 165–173. https://doi.org/10.1300/J018v05n01_09

Sachdev, P. S., Brodaty, H., Reppermund, S., Kochan, N. A., Trollor, J. N., Draper, B., Slavin, M. J., Crawford, J., Kang, K., Broe, G. A., Mather, K. A., Lux, O., & Memory and Ageing Study Team. (2010). The Sydney Memory and Ageing Study (MAS): Methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70-90 years. *International Psychogeriatrics, 22*(8), 1248–1264. https://doi.org/10.1017/S1041610210001067

http://www.stanford.edu/~yesavage/GDS.html

Appendix D1

EduG analyses output for the total GDS-15, including observation and estimation design, ANOVA and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	
Occasion	O	5	INF	

Analysis of variance

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
P	367.345	353	1.041	0.012	0.012	0.012	8.6	0.001
I	56.702	14	4.050	-0.005	-0.005	-0.004	0.0	0.002
O	144.311	4	36.078	0.005	0.007	0.007	4.7	0.004
PI	496.152	4942	0.100	0.004	0.004	0.004	2.9	0.000
PO	152.542	1412	0.108	0.002	0.007	0.007	5.0	0.000
IO	679.202	56	12.129	0.034	0.034	0.034	23.6	0.006
PIO	1567.145	19768	0.079	0.079	0.079	0.079	55.1	0.001
Total	3463.398	26549					100%	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.012		.....		.....	
	.....	I	.....		(0.000)	0.0
	.....	O	.....		0.001	48.5
	.....	PI	(0.000)	0.0	(0.000)	0.0
	.....	PO	0.001	100.0	0.001	51.5
	.....	IO	.....		(0.000)	0.0
	.....	PIO	(0.000)	0.0	(0.000)	0.0
Sum of variances	0.012		0.001	100%	0.003	100%
Standard deviation	0.112		Relative SE: 0.038		Absolute SE: 0.053	
Coef_G relative	0.90					
Coef_G absolute	0.82					

Grand mean for levels used: 0.154
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.037

Appendix D2

EduG analyses output for Item 1 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	2 3 4 5 6 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.021		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.001	8.7
	.....	PI	.....		.....	
	.....	PO	0.012	100.0	0.012	91.3
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.021		0.012	100%	0.014	100%
Standard deviation	0.145		Relative SE: 0.111		Absolute SE: 0.117	
Coef_G relative	0.63					
Coef_G absolute	0.61					

Grand mean for levels used: 0.097
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.036

Appendix D3

EduG analyses output for Item 2 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 3 4 5 6 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.009		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.012	36.7
	.....	PI	.....		.....	
	.....	PO	0.021	100.0	0.021	63.3
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.009		0.021	100%	0.033	100%
Standard deviation	0.096		Relative SE: 0.145		Absolute SE: 0.182	
Coef_G relative	0.31					
Coef_G absolute	0.22					

Grand mean for levels used: 0.203
Variance error of the mean for levels used: 0.012
Standard error of the grand mean: 0.110

Appendix D4

EduG analyses output for Item 3 of the GDS-15, including observation and estimation design and G-study table

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 4 5 6 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.014		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.010	36.3
	.....	PI	.....		.....	
	.....	PO	0.017	100.0	0.017	63.7
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.014		0.017	100%	0.026	100%
Standard deviation	0.119		Relative SE: 0.129		Absolute SE: 0.162	
Coef_G relative	0.46					
Coef_G absolute	0.35					

Grand mean for levels used: 0.163
Variance error of the mean for levels used: 0.010
Standard error of the grand mean: 0.098

Appendix D5

EduG analyses output for Item 4 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 5 6 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.015		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.001	7.9
	.....	PI	.....		.....	
	.....	PO	0.012	100.0	0.012	92.1
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.015		0.012	100%	0.013	100%
Standard deviation	0.121		Relative SE: 0.111		Absolute SE: 0.115	
Coef_G relative	0.54					
Coef_G absolute	0.52					

Grand mean for levels used: 0.088
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.034

Appendix D6

EduG analyses output for Item 5 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 6 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.017		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.012	38.0
	.....	PI	.....		.....	
	.....	PO	0.019	100.0	0.019	62.0
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.017		0.019	100%	0.031	100%
Standard deviation	0.131		Relative SE: 0.139		Absolute SE: 0.177	
Coef_G relative	0.47					
Coef_G absolute	0.36					

Grand mean for levels used: 0.203
Variance error of the mean for levels used: 0.012
Standard error of the grand mean: 0.110

Appendix D7

EduG analyses output for Item 6 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 7 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.017		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.011	43.4
	.....	PI	.....		.....	
	.....	PO	0.015	100.0	0.015	56.6
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.017		0.015	100%	0.026	100%
Standard deviation	0.131		Relative SE: 0.122		Absolute SE: 0.162	
Coef_G relative	0.54					
Coef_G absolute	0.39					

Grand mean for levels used: 0.164
Variance error of the mean for levels used: 0.011
Standard error of the grand mean: 0.107

Appendix D8

EduG analyses output for Item 7 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 8 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.015		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.001	7.0
	.....	PI	.....		.....	
	.....	PO	0.013	100.0	0.013	93.0
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.015		0.013	100%	0.014	100%
Standard deviation	0.121		Relative SE: 0.116		Absolute SE: 0.120	
Coef_G relative	0.52					
Coef_G absolute	0.50					

Grand mean for levels used: 0.095
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.033

Appendix D9

EduG analyses output for Item 8 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 9 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.014		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.012	36.6
	.....	PI	.....		.....	
	.....	PO	0.021	100.0	0.021	63.4
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.014		0.021	100%	0.033	100%
Standard deviation	0.116		Relative SE: 0.145		Absolute SE: 0.182	
Coef_G relative	0.39					
Coef_G absolute	0.29					

Grand mean for levels used: 0.213
Variance error of the mean for levels used: 0.012
Standard error of the grand mean: 0.111

Appendix D10

EduG analyses output for Item 9 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 10 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.017		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.011	41.9
	.....	PI	.....		.....	
	.....	PO	0.016	100.0	0.016	58.1
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.017		0.016	100%	0.027	100%
Standard deviation	0.132		Relative SE: 0.125		Absolute SE: 0.164	
Coef_G relative	0.53					
Coef_G absolute	0.39					

Grand mean for levels used: 0.168
Variance error of the mean for levels used: 0.011
Standard error of the grand mean: 0.106

Appendix D11

EduG analyses output for Item 10 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 11 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.020		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.001	8.2
	.....	PI	.....		.....	
	.....	PO	0.012	100.0	0.012	91.8
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.020		0.012	100%	0.013	100%
Standard deviation	0.140		Relative SE: 0.108		Absolute SE: 0.112	
Coef_G relative	0.63					
Coef_G absolute	0.61					

Grand mean for levels used: 0.090
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.033

Appendix D12

EduG analyses output for Item 11 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 10 12 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.014		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.012	37.6
	.....	PI	.....		.....	
	.....	PO	0.020	100.0	0.020	62.4
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.014		0.020	100%	0.032	100%
Standard deviation	0.118		Relative SE: 0.141		Absolute SE: 0.179	
Coef_G relative	0.41					
Coef_G absolute	0.30					

Grand mean for levels used: 0.203
Variance error of the mean for levels used: 0.012
Standard error of the grand mean: 0.110

Appendix D13

EduG analyses output for Item 12 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 10 11 13 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.018		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.010	37.1
	.....	PI	.....		.....	
	.....	PO	0.016	100.0	0.016	62.9
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.018		0.016	100%	0.026	100%
Standard deviation	0.133		Relative SE: 0.128		Absolute SE: 0.161	
Coef_G relative	0.52					
Coef_G absolute	0.41					

Grand mean for levels used: 0.165
Variance error of the mean for levels used: 0.010
Standard error of the grand mean: 0.099

Appendix D14

EduG analyses output for Item 13 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 10 11 12 14 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.025		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.001	7.8
	.....	PI	.....		.....	
	.....	PO	0.012	100.0	0.012	92.2
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.025		0.012	100%	0.013	100%
Standard deviation	0.157		Relative SE: 0.110		Absolute SE: 0.115	
Coef_G relative	0.67					
Coef_G absolute	0.65					

Grand mean for levels used: 0.099
Variance error of the mean for levels used: 0.001
Standard error of the grand mean: 0.034

Appendix D15

EduG analyses output for Item 14 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 10 11 12 13 15
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.011		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.012	35.8
	.....	PI	.....		.....	
	.....	PO	0.021	100.0	0.021	64.2
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.011		0.021	100%	0.033	100%
Standard deviation	0.107		Relative SE: 0.145		Absolute SE: 0.181	
Coef_G relative	0.35					
Coef_G absolute	0.26					

Grand mean for levels used: 0.205
Variance error of the mean for levels used: 0.012
Standard error of the grand mean: 0.108

Appendix D16

EduG analyses output for Item 15 of the GDS-15, including observation and estimation design and G-study table

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	354	INF	
Item	I	15	15	1 2 3 4 5 6 7 8 9 10 11 12 13 14
Occasion	O	5	INF	

G Study Table
(Measurement design P/IO)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.019		.....		.....	
	.....	I	.....		.....	
	.....	O	.....		0.009	35.2
	.....	PI	.....		.....	
	.....	PO	0.016	100.0	0.016	64.8
	.....	IO	.....		.....	
	.....	PIO	.....		.....	
Sum of variances	0.019		0.016	100%	0.024	100%
Standard deviation	0.139		Relative SE: 0.126		Absolute SE: 0.156	
Coef_G relative	0.55					
Coef_G absolute	0.44					

Grand mean for levels used: 0.158
Variance error of the mean for levels used: 0.009
Standard error of the grand mean: 0.093

Appendix E

Supplementary table S1. *G-Theory model estimates' calculation including component variances, formula for the design of person by item by occasion, express as P x I x O.*

Indices	Characteristic (Formula)
$X =$	observed score of a person on a particular item across occasions
μ	grand mean of X
$+X_p$	person effect ($\mu_p - \mu$)
$+X_i$	item effect ($\mu_i - \mu$)
$+X_o$	occasion effect ($\mu_o - \mu$)
$+X_{pi}$	person x item effect ($\mu_{pi} - \mu_p - \mu_i + \mu$)
$+X_{po}$	person x occasion effect ($\mu_{po} - \mu_p - \mu_o + \mu$)
$+X_{io}$	item x occasion effect ($\mu_{io} - \mu_i - \mu_o + \mu$)
$+X_{pio}$	residual/person x item x occasion effect ($\mu_{pio} - \mu_{pi} - \mu_{po} - \mu_{io} + \mu_p + \mu_i + \mu_o - \mu$)
σ^2_p	person variance component ($MS_p - MS_{pi} - MS_{po} + MS_{pio}/n_i n_o$
σ^2_i	item variance component ($MS_i - MS_{pi} - MS_{io} + MS_{pio}/n_p n_o$
σ^2_o	occasion variance component ($MS_o - MS_{io} - MS_{po} + MS_{pio}/n_i n_o$
σ^2_{pi}	person x item variance component($MS_{pi} - MS_{pio}/n_o$
σ^2_{po}	person x occasion variance component ($MS_{po} - MS_{pio}/n_i$

σ^2_{io}	item x occasion variance component $(MS_{io}-MS_{pio})/n_p$
σ^2_{pio}	residual/ person x item x occasion variance component: (MS_{pio})
σ^2_δ	relative error variance ($\frac{\sigma^2_{pi}}{n_i}+\frac{\sigma^2_{po}}{n_o}+\frac{\sigma^2_{pio}}{n_i n_o}$)
σ^2_Δ	absolute error variance ($\frac{\sigma^2_0}{n_o}+\frac{\sigma^2_i}{n_i}+\frac{\sigma^2_{pi}}{n_i}+\frac{\sigma^2_{po}}{n_o}+\frac{\sigma^2_{io}}{n_i n_o}+\frac{\sigma^2_{pio}}{n_i n_o}$)
G_r	relative G-coefficient ($\frac{\sigma^2_p}{\sigma^2_p+\sigma^2_\delta}$)
G_a	absolute G-coefficient ($\frac{\sigma^2_p}{\sigma^2_p+\sigma^2_\Delta}$)
SCI	state component index ($\frac{\sigma^2_{po}}{\sigma^2_{po}+\sigma^2_p}$)
TCI	trait component index ($\frac{\sigma^2_p}{\sigma^2_{po}+\sigma^2_p}$)

Note: *MS* stands for the mean of effect square; n_i: number of items; n_o: number of occasions; n_p: number of persons/participants

References:

Benton, A. L. (1967). Problems of test construction in the field of aphasia. Cortex: A Journal Devoted to the Study of the Nervous System and Behavior.

Benton, A. L., Sivan, A. B. and Spreen, O. (1996). Der Benton Test. Bern: Huber.

Kaplan, E., Goodglass, H., Weintraub, S. and Segal, O. (2001). Boston naming test. Philadelphia: Lippincott Williams & Wilkins.

Reitan, R. M., & Wolfson, D. (1993). The Halstead–Reitan neuropsychological test battery. In: Theory and clinical interpretation. South Tucson: Neuropsychology Press.

Spreen, O. (1977). Neurosensory center comprehensive examination for aphasia. Neuropsychological Laboratory.

Strauss, E., Sherman, E. M. and Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms, and commentary: American Chemical Society.

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. Archives of clinical neuropsychology, 19(2), 203-214.

Wechsler, D. (1981). Wechsler adult intelligence scale: WAIS-R manual: Harcourt Brace Jovanovich [for] The Psychological Corporation.

Wechsler, D. (1997). Wechsler memory scale (WMS-III) (Vol. 14): Psychological Corporation San Antonio, TX.

Wechsler, D. (2003). WISC-IV: Administration and scoring manual: Psychological Corporation.