Evaluating the e-PASS: The Psychometric Properties and User Experience of an Online Clinical Diagnostic Assessment Program for Mental Disorders

David Phong Nguyen

BCom, BSc (Honours)

Submitted as partial fulfillment of the requirements for the

Doctor of Philosophy (Clinical Psychology)

Swinburne University of Technology

Faculty of Life and Social Sciences

2013

Abstract

Diagnostic assessment is critical for identifying and treating mental disorders. However, consumers face numerous barriers to effective diagnostic assessment in traditional clinical practice, including: stigma and poor mental health literacy; high costs and shortage of clinicians; and inadequate assessment resources and training for clinicians. With recent internet usage growth, web-based diagnostic assessment programs offer a means of overcoming these barriers while also providing benefits such as added accuracy and efficiency through automated item and scoring administration. However, they also raise issues which emphasise the need to ensure such programs are effective. At present, various web-based programs are publically available, although most present limited diagnostic functionality (e.g., narrow scope of disorders, limited feedback) and lack supporting research. This prompted development of the e-PASS, a freely available web-based clinical assessment program providing diagnostic and referral feedback for 21 disorders. To inform its appropriate use and address the dearth of empirical literature regarding similar applications, this thesis reports an evaluation of the e-PASS's performance in real-world circumstances. Participants consisted of Australian adults (*N* = 616, *M* age = 38.1 years) recruited online and representing actual e-PASS users. In a psychometric study, participant subsets underwent: a structured clinical interview (*n* = 158); standardised clinical questionnaires (*n* = 173); and e-PASS re-testing (*n* = 39). With the clinical interview as the gold standard, the e-PASS's diagnostic results displayed: strong agreement for some disorders (e.g. panic disorder) but not others (e.g. social phobia, OCD); high specificity, and varying sensitivity; similar or better accuracy compared to certain clinical questionnaires; and limited accuracy in primary diagnosis. Test-retest reliability was found to be high, while logistic regression indicated that reducing the diagnostic threshold and adding particular items could improve e-PASS validity. Two studies evaluated the e-PASS's user experience, with participants (recruited from the psychometric study) completing an online survey (*n* = 88) or in-depth qualitative interviewing (*n* = 15). Results showed high e-PASS acceptability, including motivation to re-take the e-PASS for future needs. The e-PASS was widely praised for being more comprehensive and less-judgmental than previous assessment experiences, promoting accessibility, anonymity, self-reflection, and disclosure. Group differences in experiences, as well as criticisms regarding length, restrictive response options, and lack of support in some cases highlighted the need to tailor online assessment to optimise outcomes. In conclusion, the e-PASS clearly demonstrates the potential to enhance traditional clinical assessment practice, and improve consumer mental health awareness and access to appropriate services.

Acknowledgements

I am indebted to my supervisors for their support throughout my research. To Britt Klein and David Austin, thank you for welcoming me into the eTherapy unit and for fostering my research. To Britt, in particular, seeing your hard work and vision has been awe-inspiring. To Denny Meyer, your kindness, care, and ongoing support kept me on track and always feeling positive – for this, I cannot thank you enough. To Jo Abbott, it's been a pleasure working with you and I thank you for joining as a supervisor and your generous help in the final stretch of my research.

To my classmates (Danielle, Jacq, Jim, Shelley, Simmone, and especially Suz) who have shared the rollercoaster that is postgraduate research and training at Swinburne University, your friendship and support has been invaluable and I look forward to keeping in touch and following your successful careers as clinical psychologists.

A big thank you to my parents for the care and encouragement throughout my life. I also thank my brother for helping me learn about myself and showing me the meaning of perseverance.

To my wonderful partner, Tam, for your patience, warmth, and support. I'm not sure exactly how much our relationship has sacrificed for my studies, but I look forward to making up for it.

A special thank you to Nenagh – while it's been several years since you supervised me in my Honours year, you remain an inspiration.

Thank you to the Swinburne research staff (especially Robyn), my lecturers, and placement supervisors for their assistance, teaching, and support during my studies.

Finally, I am grateful to the participants in this research for trying the e-PASS, generously contributing to this research project, and most of all, having the courage to share their experiences with mental illness.

Declaration

I declare that this thesis does not does contain material which has been accepted for the award of any other degree. To the best of my knowledge and belief, this thesis does not contain material previously published or written by another person except where due reference is made in the text. Furthermore, I declare that the preparation of this report and the underlying research was conducted in adherence to the ethical requirements as specified by the Swinburne University Human Research Ethics Committee.

Name: David Phong Nguyen

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction and Overview of the Thesis

Dealing with mental illness begins with its recognition. For clinicians, this is reflected in the emphasis on clinical assessment, and diagnosis in particular, as a crucial process in identifying and treating mental disorders. For those with mental illness, undergoing an assessment and receiving diagnosis can be a powerful experience, and one that can produce outcomes considered positive (e.g., insight and empowerment) as well as negative (e.g., stigma, costs of ineffective treatment). Given its importance and potential repercussions, diagnostic assessment has been an activity typically reserved for administration by clinicians in a formal setting, However, clinical diagnostic assessment is not a straightforward process, and in practice, is limited by a various issues that explain why mental disorders are often incorrectly and/or under diagnosed (Prins et al., 2008).

It is apparent that many people with mental illness are turning to online resources to either compliment or be used in place of traditional services such as consulting a mental health professional. In response, there is a growing number of online interventions being offered, including CBT treatment programs, information websites, support group forums, and online counseling services. However, the need to use these interventions, followed by the choice of which to access is obscured in those without a clear diagnosis of their mental health issues. Amongst the online resources available are a range of online assessment options, though there are few freely available clinical assessment instruments that have demonstrated clinical validity and utility. In particular, there is a notable absence in the availability and research of more comprehensive web-based diagnostic tools that go beyond simple symptom measures.

To help address clinical assessment needs of online consumers and clinicians, the e-PASS was developed and made available at Anxiety Online (www.anxietyonline.org.au), where it also functions as an initial assessment in referring to online treatment programs. Between the launch of Anxiety Online in October 2009, to February 2013, the e-PASS has been completed by a broad range of people from around the world over 15,000 times, a phenomenal number reflecting the e-PASS's appeal, accessibility, and scalability in terms of delivering a service on a wide-scale. However, as with any online assessment tool, and particularly one that is openly accessible and provides diagnostic and referral information, it is crucial to ensure the e-PASS is functioning as intended and to provide evidence of this (International Test Commission, 2005).

Hence, the main aim of this research was to evaluate the e-PASS, and in doing so, highlight its strengths and areas for improvement, and more generally, to address the question of

whether/how such a program could adequately perform as a diagnostic tool and meet the needs of people with mental illness. In evaluating the psychometric properties of the e-PASS, there were several forms of validity and reliability to consider. However, given practical constraints, this research focused on the e-PASS's criterion validity and test-retest reliability. This met minimum requirements in demonstrating the psychometric properties of a diagnostic instrument (Ruland, Bakken, & Røislien, 2007), though future investigation should consider other forms of validity and reliability. Nevertheless, this research is the first (to the knowledge of this author) to evaluate both the validity and reliability of a web-based diagnostic program targeting a range of disorders. In reporting the psychometric evaluation of the e-PASS, the STARD guidelines were adhered to as recommended in promoting quality research of diagnostic instruments (Meyer, 2003).

The second focus of this research concerns the user experience of the e-PASS. While optimum validity and reliability remain central to the value of a diagnostic instrument, there is also a need to consider factors related to clinical utility, such as accessibility, ease of use, acceptability, and follow-up outcomes (Newman, Rugh, & Ciarlo, 2004; Smart, 2006). In this research, these factors are broadly subsumed under the concept of user experience and are arguably of critical importance in the delivery of the e-PASS and other web-based diagnostic programs, given that these programs are likely accessed autonomously by consumers. Despite the potential significance of the user experience, no known empirical studies to date have explored how consumers perceive, undertake, and respond to web-based clinical assessment programs in detail. To address this, a largely quantitative study followed by a qualitative study was conducted to provide a thorough mixed-methodological inquiry into the e-PASS experience.

A priority of this research was to examine e-PASS performance by actual users in real-world conditions. Many studies of online clinical assessment instruments recruit participants who may not be target users of their program (e.g., university cohort, participants of clinical trials). This can have implications on reported psychometric properties as well as user experience. In contrast, this research focused on actual e-PASS users (i.e., individuals who would have otherwise completed the e-PASS), and allowed them to complete it in a setting of their choice. In doing so, this design aimed to produce more ecologically valid results or, in other words, results that are more relevant to actual e-PASS use (as opposed to use under more controlled conditions). This approach is consistent with stage three of the Stage Model of Behavioural Therapies Research, which focuses on the real-world implications of web-based programs (Danaher & Seeley, 2009). Prior evaluation (e.g., expert review and pilot testing) as relevant to earlier stages of the Stage Model are briefly detailed in this thesis.

This thesis is divided into seven chapters, beginning with a broad literature review, followed by three empirical studies reporting on the psychometric properties and user experience of the e-PASS, and ending with a general discussion and conclusion. Chapter One includes a brief review of the impact of mental illness, including its prevalence and correlates in terms of burden, socio-demographic disadvantage, and economic costs. This gives context to the conditions of those faced with mental illness, and highlights the need for effective identification of mental disorders. Following is a review of the role, process, and methods used in conducting clinical diagnosis and assessment of mental disorders, and a discussion of some of the issues that limit this in current practice. In particular, the chapter highlights the inaccessibility of effective clinical diagnosis in traditional services such as consulting a health professional.

In Chapter Two, the focus is on the role of computers in aiding clinical assessment and the emergence of the internet in offering various online assessment methods. Amongst these are web-based clinical questionnaires, some of which are based on existing questionnaires and interviews, and the advantages/disadavantages of these as well their empirical performance is reviewed. In Chapter Three, an overview of the e-PASS's design and development is provided, and the chapter includes a brief report of previous pilot-testing and community validation performed on the e-PASS. The chapter ends with an outline of the general aim of this research. In the first of three chapters reporting the empirical studies undertaken in this research, Chapter Four is focused on providing evidence of the e-PASS's psychometric properties, specifically regarding criterion validity and test-retest reliability. Compared to the results of a structured clinical interview and clinical scales, the e-PASS is shown to have strong diagnostic reliability and good validity for some disorders, but not others. However, logistic regression analysis suggests several revisions to improve diagnostic performance of the e-PASS for certain disorders.

Chapters Five and Six report investigations of the user experience of the e-PASS using two distinct approaches. Specifically, Chapter Five details a predominantly quantitative study involving an online survey examining how the e-PASS is rated along specific experiential factors in comparison to a clinical interview, as well as particularly liked/disliked qualities of the e-PASS. The study also considers experiential factors that could predict the e-PASS's diagnostic accuracy. Chapter Six extends on the previous chapter by detailing a qualitative exploration of the e-PASS user experience amongst a small group of individuals. A phenomenological analysis highlights the variation in experiences between individuals and identifies a range of positive and negative characteristics and outcomes associated with the e-PASS. Importantly, the results reveal significant threats to the e-PASS's validity, but also highlight instances where the e-PASS is

advantageous (e.g., in eliciting key diagnostic information) compared to traditional assessment. The seventh and final chapter summarise the results and findings of this research, discussing potential applications of the e-PASS in light of these. Areas of future research related to the e-PASS and other online forms of assessment are considered, before a final conclusion is provided. Abbreviations frequently used throughout the thesis are outlined in Appendix G.

1.0 Chapter One: Mental Illness and Clinical Assessment

1.1 The prevalence of mental disorders.

A mental disorder (or mental illness) can be broadly defined as a clinically significant set of symptoms relating to mood, thoughts, or behaviour, and associated with present or relative risk of distress and impairment in functions (e.g., work and social; American Psychiatric Association [APA], 2000). Symptom patterns underlying mental disorders differ from the normal psychological development or response to an event considered normal within a person's culture (APA, 2000). While the construct of mental disorders lacks agreement and precision (Widiger & Trull, 1991), it undoubtedly represents a wide-spread problem causing considerable impact on individuals, communities, and societies.

Approximately 450 million people worldwide are estimated to have a mental disorder (World Health Organisation [WHO], 2001). While prevalence rates appear to have stabilised in recent years (e.g., Jorm & Reavley, 2012; Kessler et al., 2005; Spiers et al., 2011), mental disorders continue to affect a large proportion of the population. In Australia, a nation-wide survey found that 45.5% of Australians aged 16 to 85 years (7.3 million people) experienced mental illness during their lifetime, while approximately one-in-five adult Australians experienced a mental disorder in the past 12 months (Slade et al., 2009). Furthermore, the survey found anxiety disorders were the most common mental disorders, affecting 14.4% of Australians adults during a 12 month period (Slade et al., 2009).

In addition, post-traumatic stress disorder (PTSD; 6.4%) had the highest prevalence within the anxiety disorders, followed by social phobia (4.7%), agoraphobia (2.8%), and generalised anxiety disorder (GAD; 2.7%). The co-morbidity or co-occurrence of mental disorders is also apparent, with one in four Australians with mental illness experiencing two or more classes of mental disorder (Slade et al., 2009). The most common co-morbidity identified was that of anxiety disorders and affective disorders (e.g., major depressive disorder [MDD]), while a third of Australians with a mental disorder were found to also have a co-morbid physical condition (Slade et al., 2009). These results resemble international prevalence rates which, although vary between countries, consistently show anxiety (followed by mood) disorders as the most prevalent mental disorders (Somers, Goldner, Waraich, & Hsu, 2006).

1.2 The impact of mental disorders.

A global population study by WHO (2004) examined the burden caused by mental disorders amongst other conditions, and found mental disorders were amongst the leading causes of disease burden (measured by mortality risk and number of years of "healthy" life lost). For example, unipolar depressive disorders had the third highest level of burden worldwide, and in middle and high-income earning countries, accounted for the most burden of disease (WHO, 2004). Similarly, mental disorders are the leading cause (24%) of non-fatal burden of disease in Australia (Begg, Vos, Barker, Stanley, & Lopez, 2008), with highest burden attributable to anxiety, depression, and alcohol abuse (Australian Institute of Health and Welfare, 2008). Generally, mental disorders are projected to become the leading cause of burden globally by 2030 (Mathers & Loncar, 2006; WHO, 2004).

A major contributor to the burden of mental disorders is increased mortality due to physical health problems. There is substantial co-morbidity between mental disorders and physical illness (e.g., cardiovascular diseases, diabetes, respiratory diseases; Mauer, 2006), with the presence of one increasing the risk of developing the other (Slade et al., 2009). Health risk factors associated with mental illness include higher rates of smoking, lower rates of exercise, obesity, substance use, and treatment avoidance (Mauer, 2006). Consequently, people with serious mental illness are estimated to live an average 25 years less than the general population (Lawrence, Holman, Jablensky, & Hobbs, 2003; Mauer, 2006).

The relationship between mental disorders and mortality is especially pronounced in terms of sudden and premature death. Mental disorders have been estimated to be a contributing factor in over 90% of suicide cases in the US (Moscicki, 2001), which in 2007 reached an annual figure of 34,598 (Centre for Disease Control and Enquiry, 2012). In Australia, the overall rate of suicide is approximately 10 deaths per 100,000 people, ranking suicide as the 15th leading cause of all deaths (ABS, 2012). While the causes of suicide can be complex, the presence of mental disorders is considered a major risk factor of suicidality, particular in Australian males and young adults (Galaif, Sussman, Newcomb, & Locke, 2007)

An Australian national survey indicated 5.1% of people with a single mental disorder during a 12 month period reported risk of suicide, a figure six times higher than seen in those without a disorder (Slade et al., 2009). The rate increased when individuals presented two (15.7%) or three (39.2%) classes of mental disorders (affective, anxiety, and substance use disorders) considered in the survey. While suicidal ideation has the strongest association with affective disorders (ABS, 2007), it can present amongst most mental disorders, including anxiety

disorders (Khan, Leventhal, Khan, & Brown, 2002). In general, the increased suicide risk reflects the high levels of distress and impairment inherent in mental illness.

Mental illness often arises or results in the context of social disadvantage. Studies have found a significant correlation between having a mental disorder with higher unemployment, lower levels of education, and social isolation related to stigma and discrimination (e.g. Ansseau et al., 2008; Byrne, 1999; McEvoy et al., 2011). For example, unemployed people are associated with twice the prevalence of substance use disorders and three times the prevalence of affective disorders compared to those employed (ABS, 2007). Compared to the general population, higher rates of mental illness are also seen in individuals with current or past homelessness and incarceration (ABS, 2007; Draine, Salzer, Culhane, & Hadley, 2002). The social disadvantages amongst those with mental illness highlights the additional adversities faced by this group compared to the general population.

The repercussions of mental illness also extend beyond the affected individual. Partners and family members, who are often required to provide care and support, often experience psychological issues (e.g., worrying, feelings of loss, relationship tension), and longer term consequences (e.g., impaired social and occupation functioning, financial distress, and mental illness; Baronet, 1999; Ohaeri, 2002). Children of parents with mental disorders are also more likely to display mental health problems than other children (McLaughlin et al., 2012).

There are also substantial economic costs associated with mental illness, including (see Bloom et al., 2011): healthcare related expenses (e.g., assessment, treatment, and hospitalisation); non-medical costs such as welfare payments, housing accommodation, education, and research; and employment/business related costs such as lost earnings and reduced productivity. Resultingly, the global cost of mental illness in 2010 was estimated to be $US 2.5 trillion, and projected to reach US$6.0 trillion by 2030 (Bloom et al., 2011). A major contributor to overall costs is lost employee productivity, which in Australia was estimated to cost $AUD 5.9 billion per annum (Hilton, Scuffham, Vecchio, & Whiteford, 2010). According to an Australian national survey, adult Australians with affective and anxiety disorders were absent from work an average of approximately 5 out of the previous 30 days (ABS, 2007).

There has been debate around the cost-benefit of diagnosing and treating mental disorders. A particular concern is that attempts to diagnose and treat mental illness leads to increased health care costs (Lecrubier, 2001). The issue is complicated by various factors, such as the cost of ineffective treatment and additional service use associated with identifying mental illness (Paulden, Palmer, Hewitt, & Gilbody, 2009). However, adequate and effective treatment

can lead to longer term economic benefits (Hunsley, 2003). A study by Katzelnick, Kobak, Greist, Jefferson, and Henk (1997), for example, examined treatment impact on high utilisers of medical services with depression. They found that the functional improvements resulting from treatment decreased the mean cost of medical services for individuals from $13.28 per day before treatment, to $6.75 per day during antidepressant medication. These results reflect the view that while identifying and treating mental illness can be costly in the short term, it results in longer term savings compared with neglecting mental illness (Lecruibier, 2001; Paulden et al., 2009).

1.3 Course and treatment of mental disorders.

Unlike physical conditions associated with later life stages, mental disorders have a relatively early age of onset. For example, in the Australian population, the median age of onset is approximately 19 years for an anxiety disorder (McEvoy et al., 2011). While some mental disorders can be relatively brief, many present a chronic nature. Anxiety disorders in particular have been found to have a poor rate of natural remission and high rate of recurrence (e.g. Bruce et al., 2005; Yonkers, Bruce, Dyke, & Keller, 2003). A longitudinal study found the probability of recovery (symptom free for eight consecutive weeks) over a 12 year period for individuals with GAD, panic disorder with agoraphobia, and social phobia was 58%, 48%, and 37%, respectively (Bruce et al., 2005). The likelihood of recovery and remission from an anxiety disorder is worsened when co-morbid with another anxiety disorder, an alcohol or substance dependence disorder, and especially MDD (Bruce et al., 2005). In summary, mental disorders can have an insidious course which increases the need for active and effective identification and treatment.

There have been many advances in the prevention and treatment of mental disorders in recent decades. Evidence based psychosocial interventions include various types of psychotherapy (e.g., cognitive behavioural therapy [CBT], interpersonal therapy, narrative therapy, acceptance and commitment therapy, solution-focused brief therapy, dialectical behaviour therapy etc; Australian Psychological Society [APS], 2010; Hofman & Smits, 2008; Stewart & Chambless, 2009) and peer support services such as support groups (Davidson, Chinman, Kloos, Stayner, & Tebes, 1999). Other psychosocial interventions include community-based programs promoting recovery and general wellbeing through access to stable housing and transportation; improved lifestyle; social activity; and paid or voluntary work.

Amongst these, CBT has gained wide support for its application across many mental disorders (see Hawton, Salkovskis, Kirk, & Clarke, 2010), and especially anxiety disorders (Otte, 2011). Recent meta-analyses of randomised placebo-controlled trials found CBT to be highly

efficacious and effective across anxiety disorders, with the strongest treatment effect sizes (considered “large”) for OCD and acute stress disorder; and the weakest (considered “mild”, though still significant) amongst GAD and panic disorder (Hoffman & Smits, 2008; Stewart & Chambless, 2009). A particular advantage of CBT over other treatment forms is that it can promote and facilitate relapse prevention (see Hollon, 2011). However, factors which can influence the effectiveness of CBT include quality of the therapist (Kuyken & Tsivrikos, 2009) and whether adjunctive pharmacotherapy is employed (Hollon, 2011).

Pharmacotherapy (treatment via medication use) is another major treatment option for mental disorders. While its current evidence base is broad and mixed, the research literature suggests medication can be efficacious for many disorders, particularly of severe nature (Hollon, 2011). For example, some classes of antidepressants have been recommended as first line pharmacotherapy for GAD and OCD (Koen & Stein, 2011). However, amongst concerns and criticisms of pharmacotherapy (see Kirsch et al., 2008), most psychotropic medications potentially involve adverse side-effects varying in severity and likelihood of occurrence. When treating anxiety with benzodiazepines, for instance, higher doses can impair concentration, be highly sedating, and promote physical dependence over long term use (Koen & Stein, 2011).

In summary, the treatment of mental disorders is largely facilitated through pharmacotherapy and psychosocial interventions, many of which have shown to be highly efficacious and effective on their own, and in some cases, when combined or as an adjunct to one another (Otto, Smith, & Reese 2006; Roshanaei-Moghaddam et al., 2011). However, treatment effect sizes of different treatment combinations appear to vary across disorders and severity (Kirsch et al., 2008; Roshanaei-Moghaddam et al., 2011). Furthermore, there are additional treatment consequences such as the potential adverse side-effects and the burden of treatment (e.g., cost of long term therapy).

One important implication of the heterogeneity in treatment evidence is that the correct match between treatment modality and type/severity of mental disorder appears crucial for positive outcomes. For example, in cases of MDD, antidepressants exhibits low efficacy in comparison to CBT and placebo when symptoms are mild; however, antidepressants are considered advantageous when symptoms are very severe (Kirsch et al., 2008). Selecting the appropriate treatment to maximise outcomes therefore depends on correctly diagnosing the underlying mental disorder and its severity. In brief, the appropriate management and treatment of mental illness is contingent on the quality of the assessment and diagnostic process and outcome.

1.4 Clinical assessment and diagnosis.

The first step prior to any formal treatment or intervention of a mental disorder is to perform a clinical assessment. A clinical assessment can be defined as inherently a decision-making task in which a clinician iteratively formulates and test hypotheses in order to determine the nature and identity of the underlying problem and how to respond to it (Hungsley & Mash, 2007). The content and means of collecting the information used in this decision process can vary depending on the assessment setting, the purpose of the assessment, and the discipline and training of the clinician (e.g., psychiatrist versus social worker).

For example, some assessment practices might solely focus on presenting symptoms and mental state while others require a detailed background of the problem including factors such as medical, family, occupational, and developmental history. Information might be collected directly by interviewing or observing the individual of concern, through the use of testing instruments (e.g., neurological, psychological, biological), through secondary sources such as reports by others (e.g., family, other health professionals), or examining past records (Meyer et al., 2001). The course of clinical assessment activity might be brief and one-off; or it could be longitudinal and iterative (Spitzer, 1983; Meyer et al., 2001).

While the approach to clinical assessment may differ across circumstances, the main purposes of clinical assessment in relation to clinical treatment/interventions generally fall under three categories: diagnosis (e.g., determining the presence/absence of a disorder, screening of disorders, and identifying co-morbidity and severity); case conceptualisation and treatment planning; and treatment monitoring and evaluation (Hunsley & Mash, 2010). As the focus of this research is on a diagnostic assessment tool, the following will concentrate on clinical diagnosis although the other functions of clinical assessment will also be considered.

1.4.1 Diagnosis.

In relation to clinical assessment, clinical (or psychiatric) diagnosis refers to the identification of a mental disorder marked by the presence of specific signs and symptoms (Widiger & Trull, 1991). The diagnostic process involves examining an individual and their background and applying clinical knowledge (e.g. a diagnostic framework) to determine whether there is sufficient evidence to justify the presence of a disorder; if so, a diagnosis is applied. In addition, clinical diagnosis can involve gauging disorder severity and recognising co-morbidity (Hunsley & Mash, 2010). Importantly, a diagnosis represents a measure or best estimate of a disorder which can be subject to various forms of error (Kraemer, 2007), as discussed later in this

chapter. While clinical diagnosis dates many centuries, it was not until the 20th century when diagnostic classification systems were widely adopted, thus contributing to the standardisation of diagnosing mental disorders.

At present, due to the relative infancy of psychiatric knowledge and issues in nosology (e.g. validity of diagnostic systems; Widiger & Trull, 1991), the diagnostic process lacks accurate and objective neurobiological tests for distinct symptoms ("biomarkers") as utilised in many fields of medicine (Alarcon, 2009). Recently, there have been promising advances such as in the understanding of neuro-structural correlates of specific disorders (e.g. depression, bipolar disorder, schizophrenia), however, the evidence to date indicates limited diagnostic potential as no definitive biomarker has been identified (Costafreda, Chu, Ashburner, & Fu, 2009; Costafreda et al. 2011). Hence, diagnosis continues to rely on certain methods and instruments based on self-reported and observable cognitive, emotional, and behavioural symptoms, and some of these will be reviewed later in this section.

1.4.2 The purpose of diagnosis.

In modern clinical and research settings, a diagnosis can serve various purposes. More generally, it provides a commonly understood conceptualisation of a mental disorder and how it can be treated. For researchers, it efficiently categorises psychiatric problems in epidemiological studies and treatment trials (Alarcon, 2009). Classifying mental illness has helped attribute potential factors that influence the presentation, course, and outcome of mental disorders. For clinicians, a diagnosis can help label a disorder and inform appropriate resources (First & Westen, 2007). The current treatment literature is largely organised around diagnostic categories. Hence, once a clinician reaches diagnosis, they can find out relevant research (psychopathology, epidemiology, prognosis and treatment options) to assist treatment management (First, 2010; Hunsley & Mash, 2010). Diagnosis can also facilitate communication with clients for education purposes, or with other professionals for reporting and consultation. In health systems, diagnostic information can be the basis of policy-making, guide service delivery, and facilitate administrative activities such as record keeping and payment/reimbursement (Alarcon, 2009; APA, 2000).

1.4.3 Diagnostic classification systems.

Inherent in diagnosis is a process of differentiating disorders from one another and non-disorders for the purpose of better understanding different conditions and determining who might benefit from treatment (First, 2010). Diagnostic classification systems help inform this process by

providing efficient descriptions which frame the phenomenological variation in mental problems (Hungsley & Mash, 2010; Widiger & Trull, 1991). While the categorisation of mental illness dates back to the late 18th century (Kendler, 2009), it was not until the early 20th century that diagnostic categories and criteria were proposed through the pioneering work of Emil Kraepelin, followed by wide-spread standardisation and formal use across different settings (Ebert & Bär, 2010). Specifically, several diagnostic frameworks, such as the Feigner criteria (or Research Diagnostic Criteria; Feighner et al., 1972) and Diagnostic Statistical Manual (DSM; Spitzer, Endicott, & Robins, 1975) emerged and initially vied against each other in terms of validity.

Currently, however, there are two widely-used systems for distinguishing mental disorders: the *Diagnostic Statistical Manual of Mental Disorders – Fourth Edition – Text Revision* (from here on referred to as DSM-IV) produced by the APA (2000); and the *International Classification of Diseases-10, Chapter V: Mental and behavioural disorders* (ICD-10) module, produced by the WHO (2012). Other less known and employed systems include the Chinese Classification of Mental Disorders – Third Edition, published by the Chinese Society of Psychiatry, which resembles the ICD-10 and DSM-IV but has several culturally-related adjustments; and the Psychodynamic Diagnostic Manual, which draws on classical psychoanalytic theory and has a greater focus on personality functioning in explaining psychopathology (PDM Taskforce, 2006).

Most of these systems share broad characteristics: they adopt a largely categorical approach to diagnosis, where disorders and criteria are indicated by combinations of symptoms that are either present or absent in the individual of question (First & Westen, 2007); diagnostic categories are polythetic which means that different symptom combinations can indicate the same diagnosis (APA, 2000); and listed symptoms are generally not gradated, though there are subtypes of severity within a diagnostic category (e.g. mild/moderate/severe subtypes for MDD; APA, 2000). On the other hand, there are differences (e.g., wording, categorisation of symptoms) between systems which can result in minor to major variation in diagnostic outcomes. For example, despite the convergence of ICD-10 and DSM-IV criteria, the diagnostic concordance between the two systems for any mental disorder has been found to be only 68% in a population sample (Andrews, Slade, & Peters, 1999).

Although ICD-10 is the official classification system in many countries, the DSM-IV is arguably more popular in professional clinical practice and research contexts (Andrews et al., 1999). The DSM is currently being updated, with DSM-5 scheduled for release in May 2013 (www.dsm5.org, 2012). The revision process, which began in 1999, has provided an opportunity to critique major issues with the DSM and potentially address them in DSM-5 (see Philips et al.,

2012). So far, a number of changes, ranging from the wording of criteria, to the inclusion of dimensional rating and new categories of disorders, have been proposed and are undergoing review and field trials (www.dsm5.org, 2012). For example, proposed changes to the current anxiety disorder categories and criteria appear relatively minor, and include (www.dsm5.org, 2012): agoraphobia becoming a separate codable diagnosis rather than being seen to occur within the context of panic disorder; listing a panic attack as a specifier for relevant disorders, rather than as a codable disorder; separating OCD and including it in a Obsessive-Compulsive and Related Disorders category (including body dysmoprhic disorder, and hoarding disorder). However, there are strong disagreements about the proposed changes which reflect the complexity of psychiatric classification and diagnosis (Frances, 2009; Philips et al., 2012).

1.4.4 Case conceptualisation and treatment monitoring.

The mental health literature details countless concepts (cognitive, personality, interpersonal, biological, and social) that purportedly influence the cause, maintenance, and alleviation of clinical problems, however, many of these are not reflected in established diagnostic criteria (Sim, Gwee, & Bateman, 2005). As DSM-IV states in its introduction, "making a DSM-IV diagnosis is only the first step in a comprehensive evaluation" and "to formulate an adequate treatment plan, the clinician will invariably require considerable additional information about the person being evaluated beyond that required to make a DSM-IV diagnosis" (p. xxxiv-xxxv). Accordingly, clinical assessment also aims to recognise the interplay between underlying factors of a clinical problem through case conceptualisation (or "formulation").

Sim et al. (2005) reviewed definitions of case conceptualisation and summarised it as "a succinct description of the chief features of a case, as well as an encapsulation of the diagnosis, etiology, treatment options, and prognosis of a patients' problem" (p. 290). Developing a conceptualisation relies heavily on clinician judgment (Wood, Garb, Lilienfeld, & Nezworski, 2002) and involves formulating hypotheses about the causal (i.e. predisposing, precipitating, maintaining) factors of a condition. Its quality is reflected by its basis in empirically supported theory, and its acceptance by the client and other clinicians (Beiling & Kuyken, 2003). A benefit of a valid case conceptualisation is that it assists treatment planning (Hunsley & Mash, 2010), while poor treatment outcomes indicate the need to review and revise case conceptualisation. Although a diagnosis can be considered a concise conceptualisation of mental disorders, it may overlook subtle individual differences and complex factors that influence the presentation problem. As

Taylor, Abramowitz, and McKay (2010) warn, diagnostic evaluation provides valuable information but can be insufficient on its own to provide a full assessment and result in positive outcomes.

Clinical assessment can also be performed regularly to monitor changes in the presenting problem and treatment progress. This can be vital for mental disorders presenting rapid changes in symptoms and associated risk (e.g., manic symptoms in depressed individuals prompted by antidepressant medication; Thase & Sachs, 2000). Treatment monitoring can also help measure whether intermediate treatment goals are being met, and whether to change treatment or address contributing factors (e.g., client motivation, interfering behaviours; Hunsley & Mash, 2010). For these purposes, an assessment tool needs to be sensitive to change, compare outcomes with pre-treatment levels, and be sufficiently practical for repeated use (e.g., inexpensive, easy, and quick administration; Hunsley & Mash, 2010). Hence, instruments such as self-monitoring forms and symptom rating scales can be ideal for treatment monitoring and evaluation (Hunsley & Mash, 2007), and are thus often employed in treatment trials.

1.5 Evaluating a diagnostic assessment instrument.

Major changes in healthcare systems have influenced the practice of clinical diagnosis and assessment. In the US, for example, the shift from fee-for-service to a managed care model of delivering health services has created greater constraints by reducing time for assessment and the use of intensive tests (Cashel, 2002; Piotrowski, Belter, & Keller, 1998). In Australia, the recent implementation of the Better Access to Mental Health Initiative has increased access to mental health services by providing financial rebates to consumers (Pirkis, Harris, Hall, & Ftanou, 2011). However, there is now added pressure on General Practitioners (GPs) acting as "gatekeepers" to adequately identify mental illness and refer patients to appropriate services. Due to limited rebates, there is also pressure for mental health professionals to complete assessment and treatment in as brief a period as possible.

Resultingly, there is increasing need for more efficient clinical assessment means. In accordance with the move towards evidence-based health care practices in the past two decades, the use of evidence-based assessment has also been promoted in both clinical and research settings (Hunsley & Mash, 2010). According to Hungsley and Mash (2007), evidence-based assessment relies on theoretical and empirical evidence to guide the selection of constructs, the methods and measures of use, and the manner in which the assessment process is conducted for a specific purpose. However, the issue arises as to what factors should be prioritised when selecting a clinical assessment instrument for a given situation. In the

assessment literature, psychometric properties are often of main concern, though other factors are also considered important (Groth-Marnat, 2003).

A relevant concept underlying the selection of a diagnostic instrument is “clinical utility”. Clinical utility generally refers to the usefulness of an intervention or instrument and is a broadly used concept (Smart, 2006). In Smart’s (2006) multi-dimensional model of clinical utility, the core components are: appropriateness (e.g., effectiveness and relevance); accessibility (e.g., availability, cost); practicality (e.g., functionality, suitability, required training); and acceptability (e.g., to the clinician, client, and society). Others have highlighted similar factors as relevant to the choice of an assessment instrument. Pinninti et al. (2003) for example emphasised feasibility (e.g., assessment length) and acceptance (e.g., degree of ambiguous or complex items, type of response format) while Newman et al.’s (2004) guidelines refer to procedural factors (e.g., ease of administration), “clinical utility” factors (e.g., ease of interpretation), and cost factors (e.g., whether regular administration is affordable).

To develop an evidence base for an assessment instrument, Hunsley and Mash (2007) suggest evaluating the accuracy and usefulness of the assessment task, while taking into account potential errors and biases in data collection and interpretation; the various costs and benefits associated with the assessment process; and its impact on clinical outcomes for the person assessed. As different assessment methods and instruments vary along these factors, some will be more suitable for certain assessment situations over others. Importantly, when choosing an assessment instrument, one must be aware of what information is required and how it will be used in the situation (Bufka & Camp, 2010). The following sections will review popular assessment methods and instruments used for identifying mental disorders, with a focus on their psychometric characteristics as well reference important clinical utility factors such as their appropriateness, practicality, and acceptability.

1.6 Psychometrics of clinical assessment instruments.

When considering the suitability of an instrument for clinical assessment, its psychometric properties are often of most importance from a clinician’s perspective (Meyer et al., 2001; Summerfeldt, Klooserman, & Antony, 2010). The two main psychometric areas of interest fall under the categories of reliability and validity (Ayearst & Bagby, 2010).

1.6.1 Reliability.

Reliability represents the degree to which an instrument produces the same result on repeated trials, across different time points, observers, and/or samples (Golafshani, 2003). An instrument with higher reliability is less likely influenced by random error (Wasserman & Bracken, 2003). Test-retest is an important type of reliability pertaining to diagnostic assessment and refers to the consistency of diagnostic results over two separate administrations of the instrument to the same participants (Ayearst & Bagby, 2010). Inter-rater reliability is another reliability measure relevant to instruments involving interviewer judgment. To examine inter-rater reliability, the diagnostic results of two or more independent interviewers referring to identical material are compared (Summerfeldt et al., 2010). Internal reliability reflects the extent to which items in a given instrument, domain, or sub-test "hang together" and is often measured by coefficient alpha (Cicchetti, 1994). While measured separately, reliability is deemed a pre-requisite of validity in that an instrument can only be wholly valid if it demonstrates adequate reliability. However, an instrument can be reliable in producing consistent results, though not necessarily valid. Thus, reliability is considered a ceiling of an instrument's validity (Kraemer, 2007).

1.6.2 Validity.

Validity refers to the extent of which an instrument measures what it is intended to measure (Ruland et al., 2007). An instrument is considered robust in validity if there is strong evidence and theory which supports the appropriateness, meaningfulness, and usefulness of test results and interpretation (see Ayearst & Bagby, 2010). The three main forms of validity are content, construct, and criterion-related validity. Content validity refers to how well a measure represents a construct and can be assessed by carefully reviewing its content (e.g., items, scoring rules, results, interpretations) against the literature and expert opinion (Ruland et al., 2007). Construct validity refers to a measure's performance in accordance with the concepts or constructs being measured. Evaluating this could involve factor analysis, convergent validation, discriminant validation, amongst other methods (Ayearst & Bagby, 2010).

Criterion validity is the most commonly applied form of validity in diagnostic assessment literature and refers to the amount of association between a measure and a criterion. In diagnostic assessment, the criterion is usually a "gold-standard" instrument, which itself is considered a highly valid measure of the diagnostic construct in question, but is not necessarily infallible (Summerfeldt et al., 2010). Criterion validity is indicated by the association between the measure and it's criterion at a similar time (i.e., concurrent validity), or by how well the measure

predicts the criterion outcome at a later point (i.e., predictive validity; Chronbach & Meehl, 1955). Procedural validity is a distinct type of criterion validity denoting the congruence between the results of a measure and the criterion, where both share underlying operational criteria of the construct (Summerfeldt et al., 2010). For example, the procedural validity of the Mini International Neuropsychiatric Interview (MINI) schedule has been measured against a concurrent administration of the Structured Clinical Interview for DSM Disorders-IV (SCID-IV; i.e., the criterion), as both reference the same diagnostic criteria (Sheehan et al., 1995). Criterion validity is often of interest when evaluating a diagnostic instrument because the aim of the instrument is to generate similar results to the criterion. In doing so, criterion validity avoids issues underlying the theoretical construct of a disorder (i.e. construct validity; Malgady, Rogler, & Tryon, 1992).

1.7 Clinical assessment/diagnostic instruments.

There are a number of activities that could be conducted as part of clinical diagnostic assessment, including behavioural observation, neuropsychological tests, projective tests, and biological tests (Butcher, 2006; Meyer et al., 2001). However, many are not commonly used in current everyday clinical practice and research for various reasons (e.g., lack of empirical evidence of validity, reliability, practicality, and relevance). For example, projective techniques (e.g. Rorschach test) previously dominant in clinical assessment practices (and continue to exhibit utility as a personality measure and for decision-making; Butcher, 2006), have since been found to display poor inter-rater reliability and little relationship with psychiatric diagnoses, leading many to recommend against their use as diagnostic tools (Woods et al., 2002). As previously mentioned, the absence of clear biomarkers for mental disorders has also limited the validity and utility of proposed biological tests. Instead, contemporary clinical diagnostic assessment is typically based on the results of two methods, namely, clinical interviewing and self-administered questionnaires.

1.7.1 Clinical interview: Unstructured and structured diagnostic interview.

While all clinical interviewing inherently possesses a degree of structure (Groth-Marnat, 2009), it is often categorised under two distinct approaches: structured (or semi-structured) and unstructured (or open) interviews. Unstructured interviews are typically performed in everyday practice by trained and skilled clinicians and, in their purest form, do not adhere to a standardised format, assessment criteria, or questioning, as the content and procedure of the interview is dictated by the clinician (Jones, 2010). This is advantageous in allowing greater flexibility in the focus of the interview (Angle, Ellinwood, & Carroll, 1978). For example, an interviewer may prefer

to concentrate on the client's primary concerns rather than procedural matters. The unstructured format allows clinicians more room to establish rapport (e.g., empathise with clients; Groth-Marnat, 2009) and can encourage respondents to describe experiences in greater detail and depth compared with more structured interviews (Gibson, 1998).

However, the flexibility which is a particular strength of unstructured interviewer also allows a source of unreliability (Angle et al., 1978). As interviewing is conducted at the discretion of the interviewer, its quality, structure, and content are largely influenced by the interviewer's characteristics (e.g. their skills, experience, personality, theoretical perspective, and biases). During the early twentieth century, clinical diagnostic assessment was predominantly conducted via unstructured interviews. Around the 1960s and 1970s, it faced growing criticism for producing unreliable diagnoses, particularly between clinicians (Spitzer & Fleiss, 1974), and contributing to the greatly varying diagnostic rates (e.g., 11% to 50% in North America) reported in earlier research (Brugha, 1999). A desire for more reliable means of clinical diagnosis was a major factor behind the impetus to establish internationally agreed upon diagnostic criteria (e.g., DSM), which became the foundation of structured diagnostic interviews (Sheehan et al., 1998).

Structured diagnostic interviewing commonly refers to a procedure which strictly follows a standardised schedule (e.g., WHO Composite International Diagnostic Interview [CIDI]) of items in a particular sequence (Jones, 2010), and usually incorporates a rating/scoring system to reach diagnostic outcomes in accordance with a diagnostic system (Summerfeldt et al., 2010). Some schedules also employ decision rules based on previous item responses (e.g., screening questions) to, for instance, skip items deemed irrelevant. As the structured format is largely standardised, the interviewer may not require clinical expertise, though is generally required to have training in interviewing and the use of the schedule (Brugha, 1999). With some schedules, the interviewer need not make any judgments other than determining whether the respondent understood a question. This lack of pre-requisite clinical experience facilitates more cost-effective administration as needed in certain contexts (e.g., epidemiological research; Brugha, 1999).

Semi-structured interviews (e.g. Anxiety Disorders Interview Schedule-DSM-IV; ADIS-IV) vary in permitting some flexibility in how items are administered and scored. For example, an interviewer might apply clinical judgment and decision making when scoring symptom severity or asking follow-up questions to clarify symptoms (Brugha, 1999). Semi-structured interviews aim to combine the emphasis on flexibility, professional judgment, and rapport of open interviewing, with the structure of interview schedules and questionnaires (Endicott, 2001). Hence, semi-structured interviewing is usually reserved for interviewers with adequate clinical background (Brugha,

1999). Semi-structured (as with structured) interviews rely on standardised materials/procedures to reduce potential error associated with: information variance (e.g., basing diagnoses on different information); interpretation variance (i.e., interpreting the same information differently), and; criterion variance (e.g., defining disorders differently; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009).

Structured/semi-structured interview schedules were developed for and continue to be heavily used in research contexts as they reflect researchers' priorities of being comprehensive, detailed, and precise (Sheehan et al., 1998). For example, the SCID–IV has many items covering numerous disorders and pertinent clinical information (Endicott, 2001). With built-in decision rules closely adhering to DSM-IV criteria, the SCID-IV is advantageous in addressing differential diagnoses and identifying co-morbidity. Evaluation of the SCID-IV suggests it produces comparable diagnostic results to best-estimates, and has superior validity and reliability compared with unstructured clinical interviews (Bosco et al., 2000). Hence, along with other (semi-) structured schedules, the SCID-IV is often used as the "gold-standard" when gauging criterion validity of less rigorous diagnostic instruments (e.g., Carlbring et al., 2002).

There is substantial empirical evidence showing divergent diagnostic results between unstructured and structured interviewing (Rettew et al., 2009). A meta-analysis of studies examining diagnostic agreement between clinician diagnosis (made through largely unstructured interviewing means) and structured diagnostic interviews found that agreement, as measured by the kappa statistic, was on average .27 (considered "poor") and varied between specific disorders (e.g., .27 for GAD and .86 for anorexia nervosa; Rettew et al., 2009). Results such as these are argued to highlight the psychometric strength of structured over unstructured interviews (e.g., Miller, Dasher, Collins, Griffiths, & Brown, 2001; Miller, 2001; Pinninti et al., 2003).

Miller et al. (2001), for example, compared several diagnostic interview formats in an inpatient setting. Fifty-six individuals were assessed by three clinicians using either the traditional unstructured diagnostic assessment (TDA; the standard method in the setting), the SCID-Clinical Version (SCID-CV) and the Computer-Assisted Diagnostic Interview (CADI; i.e., clinician administered interview guided by computer). The clinicians subsequently developed a consensus diagnosis based on Spitzer's (1983) "Longitudinal, Expert, All Data" (LEAD) criterion, and compared it with the results of the three interview methods to measure diagnostic accuracy. The unstructured interview results showed "fair" agreement (kappa = .43) whereas the two structured interviews displayed similarly "excellent" agreement (.81-.82). While all three methods were concluded as acceptable in diagnostic accuracy, the structured interviews were considered

significantly better. In a follow-up study examining inter-rater reliability, Miller (2001) found the TDA reliability was considered poor (kappa = .24) whilst the CADI reliability was “excellent” (.75). Differences were analysed to explain variations, and it was suggested that clinicians involved in unstructured interviewing (TDA condition) performed incomplete data collection (e.g., overlooking key diagnostic criteria), thus reducing the likelihood of diagnostic precision (Miller, 2001; 2002).

Diagnostic inaccuracy from unstructured interviews may have implications on treatment outcomes. In Miller’s (2001) study referred to above, the mean average days of inpatient stay was longer for individuals assessed by an unstructured interview (12.5) compared with a structured interview (7.7). Miller suggested the structured format informed more accurate treatment plans resulting in faster recovery. Jensen-Doss and Weisz (2008) similarly compared clinician diagnoses (based on unstructured interviewing) and researcher diagnoses (based on a structured clinical interview schedule) with the treatment of 175 youths at a mental health clinic. They found that when clinician diagnoses matched researcher diagnoses, this resulted in: better therapy engagement in terms of greater attendance and likelihood of completion; and larger reductions in parent-reported problems (Jensen-Doss & Weisz, 2008). In summary, structured interviewing appears to improve not only assessment quality, but also treatment planning and outcomes.

However, structured interviewing can have several impracticalities. In non-research settings, some schedules require a fee for use which is costly over frequent administration. Schedules can also be complex and cumbersome to administer, requiring training and a degree of clinical experience (Aboraya, France, Young, Curci, & LePage, 2005). As a result, differences in interviewer training can reduce inter-rater reliability (Kobak, 1996). Furthermore, administration times can also be lengthy; for example, the SCID-IV takes an average 90 minutes for full administration (First, William, Spitzer, & Gibbon, 2007). Printed materials can also be inconvenient to store and environmental unsustainable.

Given these limitations, it is unsurprising that comprehensive structured interviews such as the SCID-IV are rarely used outside of research contexts (Pinninti et al., 2003). In mental health and community counselling settings, for example, unstructured interviews remain the primary form of diagnosing mental disorders, despite their psychometric issues (Jones, 2010). Recognising the need for more practical clinical diagnostic instruments, briefer schedules such as the MINI (Sheehan et al., 1998) and the PRIME-MD (Spitzer et al., 1994) have been developed and evaluated, demonstrating at least good diagnostic agreement with more comprehensive schedules, and taking considerably less time to administer (e.g., the MINI takes an average of 21 minutes; Lecruibier et al., 1997).

Although structured and unstructured interviews have been distinguished, it is worth noting that unstructured interviewing often exhibit a degree of structure (Groth-Marnat, 2009). As Jones (2010) advises, unstructured interviews conducted in counselling contexts can be guided by a standard format covering domains such as identifying information, presenting program, history of presenting problems, relationship history and others. The traditional psychiatric interview is also often heavily structured, by incorporating tasks such as exploring for present and past psychiatric symptoms and conducting a mental state examination (Miller et al., 2001). With the dominance of certain classification systems and related literature, clinicians have some level of awareness of diagnostic criteria and other considerations to explore during an interview, while many more experienced clinicians would be able to closely reference schedules such as the SCID-IV, without directly using them.

Also, it should be noted that involving an interviewer during the assessment process can have several benefits not yet mentioned (see Bowling, 2005). Interviewers with a friendly motivating attitude can increase motivation and response rates, particularly with longer or less interesting schedules. They can also probe and clarify responses which can be beneficial when respondents give brief or vague responses, or when items are misinterpreted or misunderstood. Interviewers can also be trained to take into account additional factors which could influence participation, such as distracting environmental stimulus (e.g., outside noise) and intoxication. If the interviewer is a clinician with the potential to offer therapy, the interview process can also commence the process of building rapport and therapeutic alliance as a foundation for therapy.

1.7.2 Self-administered questionnaires.

Self-administered questionnaires are another common aid for clinical diagnostic assessment. Questionnaire content typically reflects diagnostic criteria or the underlying disorder construct, while item format is usually closed, with item response being dichotomous or rated on a Likert scale to facilitate scoring and research (e.g., evaluation, normalisation data). Unlike clinical interviews, the clinician's role is minimal in the response process, as respondents work independently through the questionnaire, potentially without supervision (e.g. at home via post; Joiner & Pettit, 2008). Self-administered questionnaires are advantageous over interviews as they are usually less costly and time consuming due to less clinical staffing to administer and score (Bufka & Camp, 2010). Exceptions include comprehensive measures (e.g., MMPI-2) requiring intensive scoring and interpretation. Nevertheless, self-administered questionnaires are widely

recommended for addressing clinical assessment needs in settings with time and clinician constraints (e.g. primary care; Dozois & Dobson, 2010; McAlpine & Wilson, 2004)

The literature details countless self-administered questionnaires varying in factors such as: length (e.g., several versus hundreds of items); breadth (e.g., specific symptoms, groups of disorders); purpose (e.g., symptom tracking, diagnostic screening); user-friendliness; and overall utility for certain settings (Bufka & Camp, 2010). However, despite the abundance of questionnaires, only a small proportion are widely accepted, researched, and used in practice. For example, amongst roughly 280 measures of depression available, a select few (e.g., CES-D, BDI-II) appear in the majority of recent adult depression studies (Santor, Gregus, & Welch, 2006).

In addition to measuring symptom severity and treatment response, some self-administered questionnaires are used to assist diagnosis by assigning a threshold or cutoff score which divides the scoring range into diagnostic groups (Ayearst & Bagby, 2010; Roberts & Illardi, 2008). For example, the Social Phobia Inventory (SPIN) targets social phobia with a total score of 0-68 based on the 0-4 ratings of 17 items (Connor et al., 2000). While higher SPIN scores reflect more severe or numerous symptoms, empirical evidence suggests a score of 19 or above distinguishes individuals with social phobia with reasonably good accuracy (.73 sensitivity, .84 specificity; Connor et al., 2000).

As with other clinical assessment instruments, an understanding of the psychometric properties of self-administered questionnaires allows an appreciation of their limitations and suitability for a particular purpose. For example, few self-administered questionnaires demonstrate sufficient diagnostic validity to warrant their use as an absolute diagnostic tool. Ayearst and Bagby (2010) raise the example of the Beck Depression Inventory-II (BDI-II), a measure of depression symptoms. The BDI-II has shown a sensitivity of .95 for clinical depression, indicating it correctly identifies 95% of depressed individuals as clinically depressed (Arnau et al., 2001). However, the BDI-II has also demonstrated a positive predictive value of .50, indicating that only 50% of those identified as having clinical depression actually have clinical depression, while the remaining 50% are “false-positive” cases (Ayearst & Bagby, 2010).

With these attributes, questionnaires such as the BDI-II are more commonly employed for screening purposes (Bufka & Camp, 2010). The aim of screening is to broadly identify positive cases (i.e., presence of a disorder) as efficiently as possible across many people (Streiner, 2003). Further to having practical benefits (e.g., low administration cost), questionnaires make valuable screening tools because their threshold score can be adjusted to maximise sensitivity so that “true positive” cases (individuals with an actual diagnosis) are not missed (Bufka & Camp, 2010).

However, higher sensitivity is usually achieved at the expense of lowered specificity and therefore higher false-positive results (Ayearst & Bagby, 2010). Therefore, when adjusting an instrument's diagnostic threshold, researchers are affecting both potential sensitivity and specificity. For many questionnaires, assigned threshold scores at best associate with moderate sensitivity and specificity, or higher levels of sensitivity with relatively low specificity (Ayearst & Bagby, 2010).

A further psychometric issue with threshold scores is whether they are suitable across different population groups and settings (Ayearst & Bagby, 2010). Generally recommended thresholds are typically calculated from sample data lacking representation of the wider population. However, a questionnaire's psychometric properties may vary with different populations. The factor structure of the BDI-II, for example, has shown inconsistency between different samples (e.g., college students, psychiatric patients; Arnau, Meagher, Norris, & Bramson, 2001). Regarding diagnostic accuracy, evidence suggests the BDI-II's generally recommended threshold score for identifying depression should be increased (from 18 to 20) when used with college students (Dozois et al., 1998).

While many studies apply thresholds to dichotomise continuous questionnaire scores into diagnostic categories, it is worth noting this may impact on treatment outcome measures. Cuijpers, Smith, Hollon, and Anderson (2010) conducted a meta-analytic comparison of psychotherapy studies where depression was measured pre- and post-testing with dichotomous and continuous outcomes. They found that the pooled effect sizes based on dichotomous and continuous outcomes were similar. However, there were noticeable differences in individual studies, with results using continuous outcomes displaying greater heterogeneity and more conservative effect sizes than dichotomous outcomes. Hence, Cuijpers et al. recommended that the two types of outcomes not be used interchangeably across time points in a situation.

There are additional limitations associated with the self-administered questionnaire format. With limited items, questionnaires may miss idiosyncratic symptoms and potentially underestimate psychopathology. As most self-administered questionnaires focus on specific symptom areas, they could also overlook co-morbidity or more severe problems when used in isolation (Dozois & Dobson, 2010). If a respondent cannot clarify items and responses, self-administered questionnaires also increase the likelihood of misinterpretation and response bias (especially if the respondent has low reading ability and language/cultural barriers; Bowling, 2005; Dozois & Dobson, 2010). Another issue affecting questionnaire and interview response in general is the influence of social expectations. Individuals have been shown to fill out questionnaires differently when amongst others as opposed to having greater anonymity (e.g., completed at

home via post; Bowling, 2005). Given these potential issues, it is often recommended that questionnaire results be followed with a more thorough assessment for clarification, rather than solely relied upon as a diagnostic source (Bufka & Camp, 2010; McAlpine & Wilson, 2004).

1.8 Client and clinician factors influencing diagnostic assessment.

Standard clinical assessment procedures rely heavily on reported information, typically from the individual under assessment. According to Kraemer et al. (2003), a respondent's report reflects several influences: the actual characteristics intended for measurement (e.g., symptoms, disorder); the respondent's circumstances; the respondent's perspective; and measurement error. In light of these factors, an important aim of assessment is to minimise the variance attributable to perspective and context (Garb, 2005). In reality, the accuracy of self-reported information is susceptible to problems such as limited insight, memory deficits, cultural biases, disordered thinking, the influence of mood, and social desirability (Dozois & Dobson, 2010; Kraemer et al., 2003). For instance, depressed patients reportedly perceive their impairment as worse than indicated by objective measures (Lahr, Beblo, & Hartje, 2007) while inconsistently report age of onset, duration, and frequency of episodes (Bromet, Dunn, Connell, Dew, & Shulberg, 1986). Hence, thorough clinical assessment is recommended to include multiple information sources (e.g. family, medical records) to corroborate an individual's reports (Meyer et al., 2001).

The quality of diagnosis also depends on the assessor's skills and experience in conducting interviews or using questionnaires. As Bhugra, Easter, Mallaris, and Gupta (2011) found in a qualitative study interviewing 31 psychiatrists, there was no uniform approach to decision making (including diagnostic decisions). Rather, approaches were influenced by clinical experience and external pressures (e.g. time constraints, available treatment) and varied between clinicians regarding the emphasis on: information gathering, intuition, evidence-based practice, cognitive reasoning, uncontrollable factors, and influence form multidisciplinary team members (Bhugra et al., 2011).

Approximately 40% of interviews conducted by trained interviewers (including physicians, psychologists, social workers, nurses) seeking professional certification reportedly fail to conduct satisfactory diagnostic assessments (Endicott, 2001). Common mistakes include missing reported cues, not clarifying symptoms when required, and not recognising inconsistent symptom profiles which indicate differential diagnostic issues (Endicott, 2001). Inevitably, clinicians exhibit information-gathering and decision making biases to some extent which affect the quality of clinical diagnosis (Garb, 2005). As Garb (2005) draws attention to, the study of heuristics and

biases suggests that clinical judgment may be prone to the same influencing factors underlying everyday decision making (e.g. “Halo” effect, confirmatory bias, affect heuristics; Groth-Marnat, 2009). For instance, an affect heuristic refers to the rapid and automatic affective response which can underlie “clinical intuition” (Garb, 2005). While this could prompt positive outcomes (e.g., conducting a risk assessment after feeling unsettled by client’s mention of hopelessness) it may also lead to negative effects such as gender and racial biases.

Another clinician bias is the tendency to emphasise symptoms more central to a clinician’s theory of disorders (Kim & Ahn, 2002). Over several experiments, Kim and Ahn (2002) found clinical psychologists made diagnostic decisions with greater weight on symptoms consistent with their own theory of disorders as opposed to assigning equal weight to relevant DSM-IV criteria. In contrast, clinicians can also overly emphasise explicit symptoms and rules as outlined in the DSM-IV, while neglecting important clinical factors such as personality (Garb, 2005). In reviewing clinical decision making, Garb (2005) concluded that clinicians rarely make judgements based on all available information. These limitations in clinical judgement help explain the low inter-rater reliability of diagnoses observed between clinicians using unstructured interviewing (Miller, 2001).

Although structured interviewing may help minimise biases, strict adherence to an interview schedule can also forego opportunities to elicit important contextual information and diminish the clinician’s responsibility to understand the presenting problem (Meyer et al., 2001). The simplicity of diagnostic criteria in representing complex phenomena makes schedules particularly susceptible to misuse by inexperienced interviewers (Endicott, 2001). For example, many schedules do not thoroughly address differential diagnoses and, therefore, inexperienced users may overlook this. Hence, Endicott (2001) argues adequate clinical skills are needed to compensate the limitations of interview schedules. Specifically, it is suggested clinicians have experience with targeted diagnoses, understand the intent of items, recognise syndromes and anomalies, and be able to clarify inconsistent or unclear responses. Overall, Endicott (2001) emphasises that the assessor is more important than the procedures used for diagnosis.

Given the various limitations of self-reported information and clinician rated information, not surprisingly differences in outcomes between the two can emerge. Cuijpers, Hoffman, and Anderson (2010) conducted a meta-analysis of 48 randomised controlled trials for depression treatment involving self-report and clinician rated instruments. They found differences in effect sizes suggesting that instruments relying on clinician ratings produced larger effect sizes than those based purely on self-report. While it was suggested that self-report instruments were either

more conservative or less sensitive to detecting change than clinician rated measures, it could not be determined whether one should be prioritised over the other for research application; hence, it was suggested that both be used in measuring outcomes (Cuijpers et al., 2010).

The issues discussed highlight the lack of diagnostic procedures or instruments in psychiatry considered optimally valid and reliable. The lack of gold-standard diagnostic assessment is particularly evident for mental health problems such as personality disorders (Garb, 2005; Meyer et al., 2001). Although structured interviews such as the SCID-IV are considered "gold-standard", they are still prone to issues that affect reliability and validity. In response to the lack of suitable criterion measures of which psychiatric diagnoses can be measured against, and the high frequency of diagnostic discrepancies in clinical practice, Spitzer (1983) proposed the LEAD standard (as earlier referred to) as a criterion for assessing procedural validity of clinical diagnostic instruments.

The "LEAD" acronym refers to longitudinal evaluation (e.g. more than one examination over different time points) performed by experts (e.g., experienced clinicians), and employing all data available (e.g., client, family, therapists, case notes, test results, medical records). Although attempts to validate the LEAD procedure suggest even the LEAD has less than perfect reliability and validity (Kranzler et al., 1995), its underlying principles of collating maximal information over time is recommended as best practice (Garb, 2005; Meyer et al., 2001). For example, Meyer et al. (2001) states that optimal clinical assessment requires "sophisticated integration of information derived from a multi-method assessment battery (p. 155)". Drawing on this, Garb (2005) suggests a diagnosis formed from ongoing therapy could be more accurate than structured interviews because it derives from extended client observation and interaction.

1.8.1 Diagnostic classification and its limitations.

In addition to measurement issues, another major factor underlying the validity and reliability of clinical diagnosis is the true nature of the phenomena in question. Research evaluating diagnostic tools such as structured interviews often focus on criterion validity, or how well they approximate other instruments following the same diagnostic classification system (Summerfeldt et al., 2010). However, an inherent issue with diagnosis is whether prevailing diagnostic systems are valid. As researchers have noted, evidence-based diagnosis requires the use of an established diagnostic criteria (Hunsley & Mash, 2010) and a diagnostic instrument is only as valid as the underlying diagnostic theory (Summerfeldt et al., 2010). While the DSM-IV is the predominant psychiatric nosology, the literature has voiced a number of serious issues (see

Phillips et al., 2012; Lemperiere, 1995, for thorough reviews) which have repercussions on the psychometrics of the instruments based on the criteria. A brief history of the DSM follows to give context to the discussion of some of these issues.

1.8.2 The DSM: A controversial diagnostic classification system.

First published in 1952, the DSM has since undergone several revisions (Wilson, 1993). The first edition reflected sociological and psychodynamic theories emphasising environmental factors in the aetiology of mental disorders (Wilson, 1993). DSM-II was released in 1968 and met growing criticism that it did not adequately distinguish normality and abnormality, nor specify symptoms to allow reliable diagnosis (Phillips et al., 2012). The need for more standardisation lead to DSM-III, published in 1980. DSM-III marked a major development by implementing an atheoretical phenomenological approach to defining mental disorders for the purposes of improving diagnostic reliability and communication (Shear, Bjelland, Beesdo, Gloster, & Wittchen, 2007). The resulting criteria were determined by "expert opinion", with less emphasis on psychoanalytic and "psychobiological" concepts, and greater emphasis on biomedical models and established operationalisations of disorders (e.g., "Scheiderian" symptoms of schizophrenia; Alarcon, 2009; Kendler, 2009).

While DSM-III exhibited face validity according to developers, it required further validation by demonstrating clear boundaries in terms of aetiology (e.g., genetic background, neurochemistry), clinical course, treatment response, functional imaging, and so on (Kraemer, 2007; Phillips et al., 2012; Robins & Guze, 1970). Rather, research in various domains has shown considerable overlap in syndromes as indicated by high diagnostic co-morbidity in epidemiological studies and the lack of treatment specificity across different diagnoses (Phillips et al., 2012). The "top-down" approach of basing diagnosis on syndromes is attributed to the limited progress in psychiatric research (e.g., etiology) and strong doubts remain as to whether DSM-IV disorders represent discrete disease entities, in the medical sense.

1.8.3 Clinical significance criterion.

A particular criticism of DSM-III was that it resulted in a high rate of false-positive diagnoses by not clearly demarcating mental disorders (Spitzer & Wakefield, 1999). To alleviate concerns of pathologising potentially normal behaviour, DSM-IV released with the inclusion of a clinical significance criterion for many diagnostic criteria sets (Spitzer & Wakefield, 1999). This stated a requirement of "clinically significant distress or impairment in social, occupational, or

other important areas of functioning" associated with symptoms of certain disorders (APA, 2000, pg. 8). Without clear definition or supporting evidence in DSM-IV text, the criterion necessitated clinical judgement to determine "significance", which contrasted with the generally prescriptive nature of DSM (Ustun & Kennedy, 2009). However, as Spitzer and Wakefield (1999) noted, narrowly applying the criterion could increase the rate of false-negative diagnoses (i.e., ruling out a disorder in fact present), while broad application could be redundant in criteria sets where distress and/or impairment are implied in symptom descriptions.

Evidence suggests the significance criterion has varying effect on diagnostic thresholds of mental disorders. Using national survey data (*N* = 10,641), Andrews, Sunderland, and Kemp (2010) compared distress and disability measures (K-10, SF-12) of people with different mental disorders and those without a physical or mental disorder. Discounting the clinical significance criterion was found to increase the prevalence of the six considered disorders from 7.2% to 9.1%. When including the significance criterion, over 90% of individuals with a dysthymic, MDD, GAD, or PTSD diagnosis had disability or distress scores above the mean score of well people. When removed, this proportion dropped to 63-74% based on the disability scale, and 65-80% based on the distress scale.

Furthermore, the proportion of individuals with social phobia or OCD scoring above the mean disability and distress score of well people was less than those with other disorders, particularly when the significance criterion was removed (Andrews et al., 2010). For example, without including the criterion, only 50% of individuals who met the remaining diagnostic criteria of social phobia had a disability score above that of well people, whilst only 33% of individuals with OCD had distress scores above that of well people. The authors concluded that diagnostic criteria for social phobia and OCD are associated with less severity than criteria defining the other mental disorders considered. Hence, they suggested increasing the diagnostic threshold of social phobia and OCD by strengthening the diagnosis related criteria, and removing the significance criterion.

The significance criterion has also been criticised for ineffectively distinguishing the underlying dysfunction resulting in symptoms of the disorder (Spitzer & Wakefield, 1999). For example, the distress and impairment accompanying mood related symptoms may result from a normal reaction to stress or loss rather than dysfunction of mood-regulating mechanisms. Making this distinction requires sound clinical judgment including the "careful assessment of the nature, course, and context of the symptoms" (Spitzer & Wakefield, 1999, pg 7), which is beyond the instruction provided by DSM-IV criteria. An implication of this is the risk of misdiagnosis when

laypersons use the DSM-IV, for example, when administering certain structured interviews without sufficient judgment to distinguish clinical significance (Brugha, 1999).

Despite these issues, it is acknowledged that gauging distress and impairment can be helpful for certain purposes, such as indicating the need for services (Moos, Nichol, & Moos, 2002; Spitzer & Wakefield, 1999). The DSM–IV has a separate scale of overall psychological, social, and occupational functioning called the Global Assessment of Functioning (GAF) which is scored from 0-100 using clinician judgement and operationalised criteria (APA, 2000). While GAF has some evidence of reliability and validity (Jones, Thornicroft, & Dunn, 1995) and is closely associated with psychiatric symptoms and diagnosis (Moos, Nichol, & Moos, 2002), the scale is often underutilised and arguably too broad as a measure for impairment (Moos, Nichol, & Moos, 2002). Interestingly, ICD-10 separates disability from the diagnosis of mental disorders by defining mental disorders as involving symptoms often associated with distress and with interference; however, it does not go so far as the DSM-IV in necessitating significant distress or impairment (Ustun & Kennedy, 2009).

1.8.4 Category versus dimensions.

Another prominent issue underlying diagnosis is the question of whether the study and practice of mental disorders would be better served by a dimensional view of mental illness as opposed to the categorical approach largely inherent in diagnostic classification systems (Kraemer, 2007; Kendall & Jablensky, 2003). According to the DSM-IV, a categorical approach is adopted across medical fields and "works best when all members of a diagnostic class are homogenous....and different classes are mutually exclusive" (APA, 2000 pg. xxxi). However, many diagnostic characteristics defined in DSM-IV can be considered dimensional in nature, as reflected in the heterogeneity of symptoms and severity when measured by scales representing the underlying constructs of a disorder (see Shear et al., 2007). Adopting a dimensional approach could better reflect the range of mental illness presented in the community, and facilitate a more gradated mental health system. It could also bypass issues of categorisation, such as the need to determine thresholds for classifying symptoms and disorders, which can be an arbitrary and subjective process and have significant implications on diagnostic outcomes (de Boeck & Wilson, 2005; Kraemer, Noda, & O'Hara, 2004).

However, a categorical approach does have practical benefits, particularly when diagnosis is used to inform a decision (Kraemer et al., 2004). For example, in practice, the treatment of mental illness requires dichotomous decisions such as whether or not to pursue a

particular treatment, and a categorical diagnosis facilitates this. Given the strengths of both approaches, a “spectrum” approach to representing mental illness, where categories are superimposed on top of dimensional constructs, is often adopted by clinicians in practice. In their review of the category versus dimension debate, Kraemer et al. (2004) earlier recommended that both be integrated in the recently released DSM-5 and adopt the following criteria: they correspond well together (e.g., highly correlated); have clinical utility (e.g., be intuitive to clinicians); and display strong psychometric properties. The push for a more dimensional view of mental illness was apparent in the development of DSM-5 (Frances, 2009), which includes the introduction of dimensional ratings for measuring: 1) personality disorders; 2) the severity of certain disorders; 3) and “cross-cutting” symptoms underlying all disorders (www.dsm5.org, 2012).

However, amongst the numerous concerns regarding the content of DSM-5 (see Phillips et al., 2012), there are doubts as to whether dimensional ratings will be accepted by practicing clinicians or improve outcomes (First, 2010; Frances, 2009). In justifying its categorical approach, the current DSM-IV-TR highlights the lack of agreement within the literature of which dimensions should be incorporated into psychiatric nosology (APA, 2000). As Shear and colleagues (2007) suggest, there are multiple dimensional approaches, such as dimensionally measuring core diagnostic features of a disorder, measuring dimensional facets common to different DSM disorders, (i.e., “cross-cutting” method); or higher order approaches which reduce disorders down to common dimensional factors (e.g., the tripartite model; Clark & Watson, 1991). The various options, each with empirical support, reflect the complexity of psychiatric classification and diagnosis.

1.8.5 Other criticisms of the DSM.

There are other notable criticisms of the DSM discussed in the literature and deserve some mentioning here. These include criticisms regarding the absence of previously dominant psychodynamic theory (Alarcon, 2009) and the tampering of classic phenomenological theory in current diagnostic criteria; as well as the absence of neuroscientific evidence in the DSM (Hyman, 2007). In addition, there are concerns about the ethnocentric (Caucasian-oriented in particular) nature of the DSM and its lack of consideration of cultural factors (Alarcon, 2009) such as norms and expectations, which may influence the presentation and perception of symptoms, quality of life, and socioeconomic correlates of mental disorders (Canino, Glorisa, Bravo, & Milagros, 1994)

Another concern which more closely ties with the personal implications of diagnosis is the effect of diagnostic labels. During the 1960's, labeling theory emerged from sociological literature controversially suggesting that labels of mental disorders could contribute to further mental illness (Scheff, 1974) as well as adverse effects in other aspects of a person's life (e.g., social and occupational functioning) due to discrimination and rejection (Link, 1982). More recent research suggests that the general public as well as health professionals react more negatively (e.g., perceive threat/danger, less sympathy and optimism, misattribution of control of symptoms) to people with diagnoses such as 'borderline personality disorder' and 'schizophrenia' compared with other disorders (Arkar, 1994; Crisp, Gelder, Rx, Meltzer, & Rowlands, 2000; Markham & Trower, 2003). Due to associated stigma and negative opinions of diagnostic labels, a diagnosis has the potential to contribute to social isolation, distress, and difficulties in functioning (Markham & Trower, 2003). This issue draws attention to the need for clinicians to conduct diagnosis as accurately and appropriately as possible, and to carefully consider diagnostic implications.

1.9 Barriers to clinical assessment and treatment.

Despite the high prevalence of mental illness and availability of evidence-based assessment and treatment, there is a relatively low uptake of mental health services in the general community (van Beljouw et al., 2010, Kessler et al., 2001; Slade et al., 2009). A US national survey, for example, found 62% of individuals with a serious mental disorder failed to receive services (Substance Abuse and Mental Health Services Administration, 2010). A similar pattern is seen in Australia where, according to the 2007 National Survey of Mental Health and Wellbeing (NSMHWB), only 35% of Australians with a mental disorder in the past 12 months reported accessing health services (Slade et al., 2009). Furthermore, services were more likely accessed by people with affective disorders (58.6%) compared with anxiety (37.8%) or substance use disorders (24%), while a greater proportion of females (40.7%) accessed services compared to males (27.5%; Slade et al., 2009).

Low service access has been associated with the lack of perceived need for care amongst consumers (e.g., Kessler et al., 2001; Slade et al. 2009), which is in part likely related to self-perceived severity of symptoms (van Beljouw et al., 2010). In Australia, these issues could be due to the limited knowledge of mental disorders (i.e., mental health literacy) amongst the general population (Coles & Coleman, 2010; Jorm, 2000). For instance, Reavley and Jorm (2011) presented vignettes of various mental illnesses to over 6000 Australians adults and asked them to identify the likely problem. Although over 75% of respondents correctly identified depression,

only a third correctly labelled PTSD, while 9.2% correctly identified social phobia. These findings suggest that despite recent overall improvements in mental health literacy (Reavley & Jorm, 2011) many people lack the ability to correctly recognise common mental disorders.

For those recognising their mental health issues, concerns about treatment inefficacy and adverse side effects could also deter consultation with a health professional (van Beljouw et al., 2010; Prins et al., 2008). Kessler et al.'s (2001) study, for instance, found that 45% of those who had a mental illness yet refused treatment expressed poor confidence in treatment efficacy to justify its use. The lack of confidence in mental health providers and treatment can also be compounded by emotional factors. In particular, stigma, embarrassment and fear of rejection can act as major barriers to service access (Prins et al., 2008). Mental illness is associated with a higher level of stigma compared to physical conditions (Corrigan & Watson, 2002) which could explain why patients usually prefer to discuss somatic symptoms over psychological or mental health symptoms in primary care settings (Bufka & Camp, 2010).

Beliefs about how mental health issues will be perceived and responded to by others can be strong determinants of help-seeking behaviour (Cepeda-Benito & Short, 1998). Reviewing factors contributing to stigma, Schulze (2007) found that individuals more likely feel stigmatised when consulting health professionals who: do not treat them as individuals; withhold information; and give diagnostic labels and limited treatment options which do not incorporate the client's background. The experience of stigma related to mental illness appears pronounced in certain population groups, such as individuals from minority backgrounds (Faye, 2005) or rural areas (Nicholson, 2008). The impact of stigma can extend beyond service access and can influence self-esteem, social inclusion, work functioning, amongst other areas (SANE, 2007).

Practical constraints such as high costs and inconvenience are considered the most significant barriers to professional help-seeking for depressive and anxiety disorders (Prins et al., 2008). In Australia, the inaccessibility of mental health services is apparent when considering the shortage of mental health professionals and cost of consultation. In 2009, the approximate number of full-time psychiatrists was 23 per 100,000 people in major cities, and 12 per 100,000 people in regional areas (AIHW, 2011). Regarding costs, the recommended assessment fee with a clinical psychologist is $AUD 222 for 46-60 minutes (APS, 2013). Although some clinicians offer discounted or rebated fees, the average net cost (including indirect costs such as travel expenses) is arguably beyond the means of many low to middle income earners. Given the practical barriers of specialist mental healthcare, it is unsurprising that only 38% of Australians with mental illness access a psychologist while 23% consult a psychiatrist (Slade et al., 2009).

For those who do access help, another barrier is the lack of adequate assessment and diagnosis in health settings, especially in primary care (Houston et al., 2001; Mitchell, Vaze, & Rao, 2009; Van Os, Van den Brink, Van der Meer, & Ormel, 2006; Weiller, Bisserbe, Boyer, Lepine, & Lecrubier, 1996). In Australia, as within many developed countries, general practitioners (GPs) take on many levels of responsibility in the primary care context, including: assessment; referral to specialised services; care coordination; and treatment. GPs are seen by over 70% of Australians with mental disorders and are usually the first point of contact when a person has a health concern (Slade et al., 2009). Furthermore, over a quarter of people with mental disorders solely access their GP for mental health assistance (Slade et al., 2009). As such, GPs are considered "gatekeepers" of mental health care in the Australian health system.

However, low and varied rates of adequate diagnosis are observed in primary care settings (van Rijswijk et al., 2009). For example, studies indicate up to 50% of cases involving MDD (Mitchell, Vaze, & Rao, 2009; Wittchen & Pittrow, 2002) and social phobia (Weiller, Bisserbe, Boyer, Lepine, & Lecrubier, 1996) are not properly identified by GPs. While evidence suggests diagnosis is more difficult in mild and brief types of disorders (Tiemens, Ormel, & Simon, 1996), even severe and persistent presentations can be undiagnosed in up to 20% of cases (Van Os et al., 2006). A recent meta-analysis found GPs correctly identified depression in approximately 47% of cases, but only recorded it in their notes in 33.6% of case (Mitchell et al., 2009). Similar findings suggest GPs are reasonably good at recognising mental problems, but relatively poor at applying diagnostic labels (Joling et al., 2011). Nevertheless, there appears to be wide variability in diagnostic accuracy between GPs (Joling et al., 2011; Mitchell et al., 2009).

The limited diagnostic assessment in primary care has been linked to several factors, including: inadequate mental health training; lack of resources (e.g., self-report questionnaires, structured interview schedules); and practical constraints (e.g., limited time) in primary care settings (Wittchen & Pittrow, 2002). Some researchers have noted the reluctance of primary care workers in exploring for mental health problems because of associated stigma, and because it is viewed as outside the core focus of primary care (Bufka & Camp, 2010; van Rijswijk et al. 2009). Others have examined attitudes of health professionals and found that on average, diagnosis was viewed as useful; however, professionals of certain disciplines (e.g., counsellors and social workers) and in particular settings (e.g., private practice) had less positive impressions of diagnostic processes and instruments than other professionals (Jensen-Doss & Hawley, 2011).

These barriers to services help explain why many people resort to self-managing their mental health (van Beljouw et al., 2010). Kessler et al. (2001) found in their population survey that

the most commonly reported reason for not seeking treatment (72%) and dropping out of treatment (58%) was to solve the problem by oneself. There are well-known benefits of self-management including greater autonomy and intrinsic motivation, and the opportunity to tailor treatment to one's need. Self-help (i.e., via resources such as books and websites) has been shown to be just as effective as face-to-face therapy (Gould & Clum, 2002), though is associated with greater rates of drop-out (Eysenbach, 2005). However, the extent of which an individual self-manages their mental illness and how effective this is could depend on many factors such as the severity of mental illness (Baillie & Rapee, 2004), the motivation of the individual, the level of adjunct support, and the efficacy of resources used (Ritterband et al., 2009).

1.10 Summary and implications.

Mental illness is a major problem with high prevalence and impact across societies. Due to the early onset and chronic nature of mental disorders, affected individuals experience prolonged distress and impairment in their functioning and, overall, have a lower quality of life than the general population. They are also at greater risk of social disadvantage and face a shorter life expectancy due to associated health conditions and higher suicide risk. The impact of mental illness extends to significant others and is responsible for major economic costs including lost productivity, costs to the health and welfare system, treatment expenses, amongst others. Treatments are available with evidence of efficacy. However, treatment response varies with specific mental disorders, the type of treatment administration, and other factors.

The role of assessment, and diagnosis in particular, is critical in recognising mental disorders and informing appropriate treatment. Advances in diagnostic classification have resulted in psychometrically supported tools to assist this process. These largely comprise self-administered questionnaires and structured clinical interview schedules, some considered gold-standard means of reaching diagnosis. However, the clinical diagnostic process has notable limitations including issues with the validity of the interview process and the classification system underlying diagnostic instruments. Nevertheless, clinical diagnosis remains integral in assessment and treatment and, when properly conducted, can result in positive outcomes.

Despite the availability of evidence based resources, many people with mental illness do not receive adequate professional care. Help-seeking behaviour and access to care can be impaired by: poor self-recognition of mental problems related to limited mental health literacy; mistrust in health professionals and treatment; the experience of stigma and other negative emotions; and practical barriers such as the lack of affordability and convenience in accessing

services. Health professionals can also add to barriers by inadequately assessing mental health problems and fuelling stigma of mental illness. Resultantly, only a portion of people with mental disorders access professional services, with lower rates particularly evident in certain groups (e.g., minority, ethnic groups; Kessler et al., 2001; Prins et al., 2008). Amongst suggestions of how these issues could be addressed, Kessler et al. (2001) highlight the need to promote self-recognition of treatment needs amongst individuals with mental illness, and for professionals, to better address patient preferences in order to increase treatment rates.

In the next chapter, the use of computers and the internet in facilitating clinical diagnostic assessment is reviewed, and it will be argued that new forms of internet-assisted clinical assessment have the potential to enhance access to accurate clinical assessment results, and in doing so, provide a range of benefits to those with mental illness.

2.0 Chapter Two: Computerised and Online Clinical Assessment

2.1. Computerised assessment.

It is first worth reviewing the role of computers in clinical assessment prior to the involvement of the internet as many of the characteristics and issues of computerised assessment extend to online assessment. The use of computers for psychological assessment dates back to the 1950s (Butcher, 2004). Computers initially scored test data but gradually incorporated data collection, test administration, and data interpretation (Butcher, 2004). Computer use for psychiatric interviewing emerged in the 1970s and was considered highly novel for both patients and clinicians (Angle et al., 1978). Interest in computerised assessment grew considerably in the 1980s (e.g., Byers, 1981; Eyde & Kowal, 1987; Roid & Gorsuch, 1984), a period when "micro" computers became more functional and accessible to the general public. Many computerised interviews and tests are still in use, some of which reflecting notable clinical instruments (e.g., MMPI-2, CIDI, SCID-IV). Such programs are usually licensed, automated, completed with or without supervision, and stored on a computer in a clinic or research setting.

2.1.1 The pros and cons of computerised assessment.

Computerised assessment offers several relative advantages compared to traditional assessment formats such as paper-pencil questionnaires and clinician interviews. Notably, the programmable nature of computers can automate processes such as item administration and scoring (Barak & English, 2002) which in turn enhances standardisation and reduce errors made by respondents (e.g., missed items) and clinicians (e.g., hand-scoring mistakes; Butler et al., 2000; Newman, Consoli, & Taylor, 1997). These benefits are especially evident when testing requires precise, complex, or time-consuming procedures (Barak & English, 2002). These qualities form part of the argument that computerised assessment can have added reliability, particularly in comparison with questionnaires or interviews administered by relatively inexperienced individuals (Butcher, 2004; Newman et al., 1997).

The automation of procedures can also save time and money, for example, when a program is largely self-administered and supervision is not required. As programs can produce results immediately, computerised assessment also offers faster feedback than traditional methods (Barak & English, 2002). Compared with paper-pencil methods, disposable materials are spared as content is stored on disk or computer drive, which over the long term has economical and environmental benefits (Barak & English, 2002). Computerised assessment also

facilitates research by making it easier to collate and store data in a format conducive for statistical analysis (Barak & English, 2002).

On the other hand, several limitations are associated with computerised assessment (Barak & English, 2002). Firstly, adequate computer hardware and software are required for administration and the initial outlay for this is costlier than traditional methods in the short term (Noyes & Garland, 2008). Computerised assessment also requires some computer literacy and skills; users require familiarity with a program's workings, particularly when novel or unintuitive (Noyes & Garland, 2008). This issue is highly pertinent when assessment outcomes depend on computer interaction (e.g., speed of a mouse click).

Furthermore, performance may be affected by individual factors such as computer experience and related anxiety (Barak & English, 2002). Reports during the 1990s indicated approximately a third of the population experienced some level of computer anxiety (e.g., computer avoidance or stressed response to use), with women and older adults more likely susceptible (Brosnan & Davidson, 1994). It was suggested that computer anxiety could invalidate computerised test results by adding unaccounted error-variance. However, findings have been inconsistent, and with rising computer literacy, it is questionable whether computer related anxiety remains a significant issue today (Noyes & Garland, 2008).

2.1.2 Is computerised assessment more reliable than clinician-administered assessment?

It is purported that computerised assessment can be more valid than human-administered assessment (Angle et al., 1978; Butcher, 2004; Newman et al., 1997). This follows the premise that computers administer items and employ decision-making processes in a more objective, systematic, and ultimately more reliable manner than humans (Butcher, 2004). A seminal paper by Meehl (1954) first concluded that assessment based on statistical and mechanical processes are more valid than clinician based decisions because of enhanced reliability. More recently, Grove, Zald, Lebow, Smith, and Nelson (2000) conducted a meta-analysis comparing clinical prediction based on clinician judgment alone, and mechanical prediction, which the authors defined as prediction based on statistics, actuarial rules, or algorithms; processes that can be programmed into a computer. The results found mechanical prediction was, on average, 10% more accurate than clinician prediction and outperformed clinician prediction in 33-47% of studies (Grove et al., 2000). The authors referred to several typical human errors in clinical judgment to explain these results: ignoring of base rates;

regression to means; heuristics (e.g., representativeness) which compromise accuracy; and lack of accuracy feedback to clinicians.

Computerised administration has also been suggested as more reliable than a clinician administered structured interviews (Angle et al., 1978). Spitzer and Endicott (1975) noted that a problem with structured interviews is that clinicians often fail to ask questions and record relevant data. Angle and colleagues (1978) explained that clinicians, despite their training, deviate from the structured process to some extent because the procedure is inherently boring, and is usually not prioritised in clinical training. In contrast, the decision-making processes of computerised assessment are pre-programmed and consistently applied across users.

Computer adaptive testing (reviewed later in this chapter) presents a good example of objective and mechanical prediction, given its computerised procedure and statistically grounded results. However, not all examples of computerised assessment clearly fall under the category of objective-based prediction. As Butcher (2004) asserts, computerised assessments may not achieve the psychometric benefits associated with objective-based predictions if output is based on untested clinical impressions rather than statistical rules supported by empirical results. However, many assessment programs produce results with some degree of subjective input on behalf of the clinician/developer. This issue highlights the need to psychometrically evaluate computerised assessment programs to ensure output is in fact valid.

2.1.3 Examples of computerised assessments.

Computerised assessment programs have largely followed the form of a standard questionnaire or test (Kobak et al., 1996). They usually begin with instructions followed by a pre-programmed sequence of items responded to by selecting an option via mouse or keyboard. The program might have built-in rules which schedule items based on previous responses. Respondents may be able to review and modify their responses, as similarly with paper-pencil questionnaires. Once all relevant items have been administered, the program would then usually compute and present results (e.g., raw/standardised score) with or without interpretative feedback, and allow data to be saved for further use (Barak & English, 2002).

Computerised assessment has appeared in several clinical domains including personality and neuropsychological assessment (see Butcher, 2004). In personality assessment, the computer administration of the MMPI/MMPI-2 has received much attention given the original instrument's popular use (Butcher, 2004). The first program to score the MMPI arose in the 1960s (Rome et al., 1962) and was followed by versions offering narrative reports derived from scale

score combinations (Fowler, 1969). The equivalence of traditional paper-pencil and computerised administration of the MMPI/MMPI-2 has been well demonstrated (see Butcher, 2004), showing little difference in scoring results (Fingers & One, 1999). Report interpretations generated by computerised MMPI-2 have also been deemed appropriate by clinicians (Butcher et al., 1998).

Computer administrations of various neuropsychological tests have also undergone examination of equivalence (Butcher, 2004). Several tests have shown comparable results across the paper-pencil and computerised-administrations (e.g., Mill Hill Vocabulary Test; French & Beaumont, 1990), while others (e.g., Standard Progressive Matrices Test; French & Beaumont, 1990) have shown discrepancies between administration types. In general, computerised neuropsychological programs have not been readily adopted by clinicians (Butcher, 2004). A major reason it seems is the lack of program sophistication and subsequent inability to amalgamate test data with important assessment factors to produce results matching that of skilled clinical judgement (Russell, 1995). However, computerised neuropsychological testing is still argued to offer screening potential, for example, in detecting early Alzheimer's, Huntington's, and Parkinson's disease (Fray, Robbins, & Sahakian, 1996).

2.1.4 Computerised clinical assessment for mental disorders.

Many programs have been developed to help assess mental disorders, some without referencing existing measures. For example, earlier programs collected psychiatric history (Carr, Ghosh, & Ancill, 1983, Ferriter, 1993) and clinic intake information (Barron, Daniels, & O'Toole, 1987). The Kraepelin comprised of 50 items and over 100 rules of reasoning to diagnose OCD (Roca-Bennasar, Garcia-Mas, Llaneras, & Blat, 1991). There have also been programs assessing alcohol related issues (e.g., Bernadt, Daniels, Blizard, & Murray, 1989; Lucas, Mullin, Luna, & McInroy, 1977). However, the majority of computerised clinical assessment programs have been based on existing questionnaires and interview schedules (Newman et al., 1997).

Programs based on self-report questionnaires imply users can autonomously undertake the program as they would complete the original (paper-pencil) measure. The conversion of a self-administered questionnaire to computerised format is considered an issue of procedural validation, with the main concern being whether the computerised version is still equivalent to the original version after changes in procedure and form (e.g., items presented on screen rather than on paper; and responded to by clicking buttons rather than by writing on paper; Kobak et al., 1996). However, adapting an instrument for computer administration may also introduce constructs to the assessment process such as computer anxiety (Schulenberg & Yutrzenka,

2001). While computer anxiety can hamper computer performance, it may also distort measures related to the experience of state based anxiety (Beckers, Wicherts, & Schmidt, 2007).

Other computerised assessment programs have been adapted from questionnaires designed to be administered by humans as a semi of fully structured interview. These include the computerised Yale-Brown Obsessive-Compulsive Scale (Y-BOCS; Rosenfeld, Anderson, Kobak, & Greist, 1992), Hamilton Depression Rating Scale (Carr, Ancill, Ghosh, & Margo, 1981; Kobak, Reynolds, Rosenfeld, Greist, 1990), Hamilton Anxiety Rating Scale (Kobak et al., 1990), and the Liebowitz Social Anxiety Scale (Katzelnick, Kobak, Greist, & Jefferson, 1995). A benefit of converting structured interviews for internet administration is that the typically large number of items and complicated branching and scoring rules often underlying interview schedules can be built into the program (Peters, Clarke, & Carroll, 1998), while certain groups might prefer the computerised format over facing an interviewer (Katzelnick et al., 1995)

However, computer administration of structured interviews raises the issue of how a program addresses the varying clinical judgment involved in interviewing (Kobak et al., 1996). For example, the interviewer process of interpreting and deciding how to classify an interviewee's response needs to be translated into a computerised process if the program is to be self-administered. A similar issue arises when converting an interview schedule into a self-rated measure. According to Newman et al. (1997), a computer cannot exhibit proper clinical judgment and must take a client's response at face value. Further to the disadvantages of computerised assessment, a computer is limited in detecting the user's needs/abilities and tailor a program (e.g., wording of items) accordingly. Hence, computerised assessment can overlook more subtle aspects of psychopathology between individuals (Newman et al., 1997). However, as with fully structured interviews which minimise input of clinical judgment, computerised versions may still provide sufficiently valid diagnostic results.

2.1.5 Computerised diagnostic interviewing.

A number of renowned diagnostic interview schedules have been converted for computer administration and shown somewhat positive psychometric results. For example, the Diagnostic Interview Schedule (DIS) is a previously popular fully structured interview schedule which can be administered by lay interviewers (Robins, Helzer, Croughan, & Ratcliff, 1981). A few studies have examined the computerised version of the DIS (Blouin, Perez, & Blouin, 1988; Erdman et al., 1992). Erdman et al. (1992) compared the diagnostic results of the computerised DIS with the standard interviewer-administered DIS, as well as a computer-prompted DIS (interviewer guided

by program). Kappa agreement statistics between the computer and interviewer administered DIS ranged from .15 to .94 across diagnoses of 20 disorders, while the average kappa between administration methods ranged from .57 to .64. These agreement levels, while mixed, were noted as similar to those observed in traditional DIS reliability studies, and it was concluded that the computerised DIS could appropriately replace an interviewer administration (Erdman et al., 1992).

The CIDI-Auto is a computerised version of the well-known CIDI interview schedule (WHO, 1993), and is intended for self-administration or to prompt lay-interviewers who then enter responses on behalf of the interviewee. The test-retest reliability of a modified CIDI-Auto when self-administered has been found to be fair to excellent (Wittchen, Lachner, & Wunderlich, 1998). The self-administered CIDI-Auto has also shown a high level of diagnostic agreement (kappa above .50) with the standard interviewer-administered CIDI (Peters et al., 1998), though others have found lower and more variable agreement rates (Peters & Andrews, 1995).

With alternative methodology, Rosenman, Korten, and Levings (1997) compared diagnoses of the self-administered CIDI-Auto against two experienced psychiatrists, and the agreement between them. According to Rosenman et al., this approach addressed the issue of whether a single psychiatrist' diagnosis is sufficiently valid/reliable to be considered "gold standard". Participants were inpatients (n = 126) who, after completing the CIDI-Auto, were then assessed by two psychiatrists unaided by an interview schedule. The CIDI-Auto made a higher average number of diagnoses than the two psychiatrists (2.3 and 1.3 respectively), whilst agreement between the CIDI-Auto and psychiatrists was significantly poorer than agreement between psychiatrists. In response, the authors proposed the psychiatrists more closely aligned in diagnosis because of their similar training backgrounds, and mutual access to patient information not factored into the CIDI-Auto. Rosenman et al. ultimately recommended against using the CIDI-Auto in place of a psychiatrist interview in clinical settings until further development and evaluation.

2.1.6 In-the-moment (in vivo) computerised assessment.

Computerised assessment programs such as those previously studied were largely administered on desktop computers in supervised clinical settings. A novel application that emerged in the 1990s involved laptops and handheld computers to facilitate more naturalistic settings for data collection (e.g., Herman & Koran, 1998; Kenardy, Fried, Kraemer, & Taylor, 1992; Taylor, Fried, & Kenardy, 1990). In addition to the general benefits of computerised assessment, the use of portable computers allowed: reporting symptoms as they occurred

throughout the day rather than retrospectively; time-stamping data entry; and prompting users to respond to questions (Herman & Koran, 1998). As such, using portable computers was regarded as advantageous for monitoring and assessing rapidly changing symptoms that can be difficult to recall (Newman et al., 1997).

In one evaluation, Herman and Koran (1998) distributed palmtop computers to thirteen outpatients with OCD. The palmtops beeped hourly to prompt completion of a modified Y-BOCS on the device. The study found only moderate agreement between the palmtop and clinician-administered Y-BOCS scores. While the small sample size limited statistic power, the results were also linked to problems with palmtop use..For example, several participants missed prompts to respond and misinterpreted the scoring scale. One participant rated low levels of OCD on the palmtop but later reported relatively severe symptoms during the clinician interview. These issues pointed to the limitations of self-administered measures as well as potential recall bias during interviewing. However, the use of portable computers for assessment provided a forerunner to online assessment in highlighting some of the strengths and weaknesses of allowing people to undertake clinical instruments with relative independence and privacy.

2.1.7 Computer adaptive testing (CAT).

CAT is another notable form of computerised assessment. With CAT, the computer uses statistical information to adapt the choice of presented items based on the respondent's ability as estimated from their response to previous items. CAT differs from conventional computerised assessment as it tailors to the individual by presenting a minimal subset of items which best measure the latent trait or characteristics of the individual (Butcher, 2004). Most clinical questionnaires and associated programs are based on classical test theory (CTT), the dominant statistical framework throughout the history of psychological testing. As such, conventional item sets are usually created with average difficulty or relevance to the latent trait in order to maximize precision amongst a group of respondents with diverse levels of the measured characteristics (Walter et al., 2007). A large set of items are usually needed to cover a wide range of trait levels with precision. Otherwise, very high or low levels of the targeted attribute are measured with less precision than more average levels.

In constrast, CAT is based on item-response theory (IRT) rather than CTT. An advantage of CAT is the ability to identify measurement precision of the target construct, based on the level of latent trait measured for the test taker (Walter et al., 2007). Rather than present a fixed set of items, CAT draws from a relatively large bank of items, each representing a relatively unique

aspect of the measured latent trait (Fries, Bruce, & Cella, 2005). CAT initially presents an item representing the most information, then selects the next item for administration based on the respondent's previous item response. Items more statistically relevant to the respondent's estimated latent trait level are subsequently presented, which spares the respondent of items deemed redundant by prior responses. The program ends once a person's trait level is measured with sufficient precision. People with trait levels closer to the population mean are usually identified with fewer items whilst those with very low or high levels are less common in the population and therefore require additional items for discrimination (Walter et al., 2007).

Overall, CAT can involve fewer items, yet produce similar levels of precision compared to corresponding conventional tests based on CTT. Although short-forms of conventional tests (e.g., screening questionnaires) can also be relatively brief, they lack the potential for high precision in CAT (Fries et al., 2005). Hence, CAT can have greater efficiency than conventional testing whilst reducing response burden and floor or ceiling effects (Butcher, 2003; Fliege et al., 2005). However, it is worth noting that CAT cannot have greater precision than that of the item bank it uses. An adaptive version of a questionnaire, for example, can only have at most the same level of accuracy as the questionnaire which presents all of the items. Thus, when time and item numbers are not constrained, the practical benefits of CAT over conventional testing are less significant (Butcher, 2003).

Interest in CAT and IRT has increased with the rise of computerised assessment, mainly in areas of educational achievement and ability testing. However, there are some examples of clinical application showing promising results (e.g. Fliege et al., 2005; Gardner et al., 2004; Walter et al., 2007). Walter et al. (2007) recently examined CAT potential for measuring anxiety in a clinical population. The study involved 2348 psychiatric and psychosomatic patients completing 13 standardised questionnaires. Of the 300 items considered, 81 were deemed relevant to the anxiety construct according to IRT. Software was developed to select, present, and score items from the 81 item pool, whilst estimating the latent trait level for each response. After simulation trials, the program was revised to 50 items and compared against common anxiety scales. The program performed well in capturing different facets of anxiety, as indicated by strong correlations with the BAI (.51), HADS-Anxiety (.66), and STAI-State (.60). As expected, the program was found to better discriminate extreme levels of the anxiety trait and have far less completion time (under three minutes) compared with the anxiety scales (Walter et al., 2007).

While these findings are promising, there are notable barriers limiting CAT implementation. Firstly, program development is very resource intensive (Walter et al., 2007). For

example, a very large sample and initial item pool is needed for selecting and calibrating the item bank, while developing the final program involves extensive computer programming and testing. Therefore, most psychometric evidence in support of CAT for clinical use has resulted from simulation data (e.g. Fliege et al., 2005; Gardner et al., 2004; Walter et al., 2007). As the CAT/IRT literature is relatively new, there also remains uncertainties regarding methodology (e.g., how best to finalise the item pool and evaluate model fit) and what effect the order of presented items has on results and response behaviour (Fleige et al., 2005).

Furthermore, there are issues regarding the nature of the constructs represented by CAT. In Walter et al.'s (2007) study, for example, it was unclear whether the program measured multiple dimensions or one underlying dimension of anxiety, and this issue was related to the choice of statistical procedure in developing the item bank. A more practical issue is the appropriateness of items for assessment needs. Many adaptive tests have constraints placed upon the item selection process (e.g., an item must be selected from a particular subset rather than the entire item bank) to ensure items meet external specifications. These constraints effectively limit the statistical benefits of CAT, though there are methods of addressing this (van der Linden & Veldkamp, 2004). Compared with conventional assessment measures, CAT/IRT is relatively novel which partly explains its lack of acceptance in the clinical community (Walter et al., 2007). Nevertheless, many researchers argue CAT has the potential to improve psychometric assessment and deserves further investigation (Fliege et al., 2005; Gardner et al., 2004).

2.1.8 Computer-driven telephone interviews.

Computers have also delivered automated telephone assessment systems and exhibited positive psychometric properties and user acceptance (Bachofen et al., 1999; Gonzalez et al., 2004). An example is the Voice-Interactive Depression Assessment System-III (VIDAS-III), a bilingual (English, Spanish) program with items based on the CES-D and DSM-IV criteria (Gonzalez et al., 2004). Users access the VIDAS-III by calling a telephone number, and are then presented digitalised voice files of assessment items. Respondents speak their reply which is digitally recognised by the system. An evaluation of the VIDAS-III found a high level of inter-item reliability (.81 to .92; Gonzalez et al., 2004). With the CIDI-SF as the validity criterion, sensitivity (.72 to 1.00) and specificity (.40 to .72) of the VIDAS-III were deemed acceptable.

High correlations (.55 to .66) between the VIDAS-III subscales and the BDI-II also supported convergent validity. Most participants rated feeling comfortable during the VIDAS-III and selected a female voice to present items. However, the study noted several issues with the

VIDAS-III (Gonzalez et al., 2004). Firstly, 12% of participant's vocal responses were incorrectly recognised which was attributed to the testing environment and the accuracy of the voice recognition software. It was also recognised that CES-D items (which VIDAS-III comprises) are not intended for diagnosis, having shown only moderate correlations with diagnostic instruments (Liang, Van Tran, Krause, & Markides, 1989, as cited in Gonzalez et al., 2004).

An accessibility issue with computerised telephone systems is the demand it places on hearing and speech, which could limit use by certain groups (e.g., people with hearing problems). However, a relative strength of computerised telephone interviewing is that consumers are likely to be familiar and comfortable with the technology, as the act of listening/responding to digitalised voice messages is commonplace. While speech recognition is imperfect, there are alternative response methods; in the BTSteps telephone assessment system, for example, respondents select responses by pressing a button on a touchtone phone (Bachofen et al., 1999). Unlike other computerised modalities, respondents can complete telephone based systems from any remote location with phone access. In this regard, computer-driven telephone interviews could be viewed as a precursor to internet-based assessments in allowing convenience and geographical distance between assessor and consumer.

2.1.9 Summary of computerised assessment.

In summary, the use of computers in facilitating clinical assessment has shown positive results across a range of applications. In particular, computers offer various benefits in the administration of questionnaires and structured interviews. Importantly, computerised versions have shown general equivalence with traditional structured clinical interviews, leading many to suggest computer administration as a feasible alternative (Blouin et al., 1988; Erdman et al., 1992; Peters et al., 1998; Levitan, Blouin, Navarro, & Hill, 1991). However, the level of diagnostic accuracy varies across programs and disorders, with some programs presenting better validity and reliability characteristics for particular disorders over others.

There were early concerns that computerised assessment would not be accepted or appropriately completed by patients, especially with severe mental health issues and limited computer experience. Although a number of studies countered this and even showed computerised assessment as preferable over a human interview, there remained resistance amongst clinicians in adopting computerised assessment perhaps due to its perceived inadequacy in comparison to traditional approaches involving clinical judgment and observation (Kobak et al., 1996; Newman et al., 1997).

While not often acknowledged, practical constraints likely also limited the proliferation of computerised assessment, especially during the 1980s and early 1990s when computers were relatively expensive, and computerised assessment involved expensive pre-loaded software normally residing on a computer in a formal setting. Specifically, computerised assessment required assessors to supply and update expensive software, computing hardware, and working space, while clients or patients were typically required to attend a clinic to undertake the assessment program. The difficulty in conducting computerised assessment from a remote location at a low cost remained a practical limitation until the internet emerged as a more cost-effective and acceptable means of providing decentralised access to computerised assessment.

2.2 The internet and its uses.

The internet is a vast de-centralised network of computers which currently pervades society. Its origins date back to the late 1950's when the US military researched means of retaining communication between computers in the advent of nuclear war (Bartram, 2002; Ryan, 2010). Advancements over subsequent decades included the development of more efficient information transfer methods within a computer network (i.e., "packet-switching") and the merging of small networks across the US (Ryan, 2010). The term "internet" first appeared in the 1970s; however, it wasn't until the 1980s that the internet resembling its current form (albeit on a smaller scale) emerged through the implementation of protocols that standardised and integrated networks around the world (Furlong, 2012).

The World Wide Web ("web") represents a portal allowing users to access and transmit information through the internet and is instrumental in every applications of the internet. The web comprises linked hypertext documents (i.e., web pages) that are navigated via a web browser. The concept behind the Web originated at the European Organisation for Nuclear Research (CERN) in the late 1980s (Ryan, 2010). By 1990, the main web components were developed including: the Hypertext Transfer Protocol (HTTP), the Hypertext Markup Language (HTML), the first web browser (The "Mosaic"), and the first HTTP servers (Bartram, 2002). In 1994, the company Netscape released what became the first popular web browser, and was soon followed by the proliferation of the web (Bartram, 2002).

Internet use has exploded over the last two decades as a result of factors such as increased affordability, improved technology, and cultural acceptance. By the end of 2011, an estimated 2.26 billion people or 33% of the world's population used the internet (Internet World Stats, 2012). In Australia, internet subscription has grown approximately 11.0% per annum in

recent years, with a total of 11.6 million internet subscribers at the end of 2011 (ABS, 2011). Internet use in Australian homes has especially thrived, with the rate of household access rising from 16% in 1998 to 64% in 2008 (ABS, 2008). The home is now the most popular location for internet use with almost all internet users having home access (ABS, 2011). While internet use continues to spread across the Australian population, there is greater access amongst higher income households, those with higher educational attainment, and younger people (ABS, 2011).

2.2.1 Internet use for mental health information.

The internet is used frequently and for various purposes. In Klein et al.'s (2010) survey of 1214 drug and alcohol website users, approximately 82% of the sample reported using the internet at least 30 minutes a day, with daily online activities including email (87.7%), social networking (57.2%), news access (48%), and general browsing (42.2%). Although reported rates vary, a large proportion of the population also use the web to source health-related information (Baker, Wagner, Singer, & Bundorf, 2003; Fox et al., 2000; Powell & Clarke, 2006). For example, a national survey indicated over half of American adult internet users had at some stage used the web to obtain health information (Fox et al., 2000). Although online searches for health information often relate to medical/physical illnesses (Fox et al., 2000; Goldner, 2006), mental health related information is also popular, with access by 10 to 39% in population surveys (Powell & Clarke, 2006; Khazaal et al., 2008).

A list of the most searched physical or mental conditions is led by depression (19%), and includes bipolar disorder (14%), anxiety disorders (9%), and sleep disorders (8%; Taylor, 1999). Although online mental health information is accessed by a range of people, sociodemographic factors associated with greater access include being female, having higher education, and being younger and of single relationship status (Baker et al., 2003; Goldner, 2006). Population data also suggests interest in searching for online information regarding mental health issues follows seasonal patterns, with increased search activity in winter months (Ayers et al., 2013). Higher income is also linked with greater access in some population surveys (Goldner, 2006), but not others (Baker et al., 2003). Web searching for mental health information is also common for people who do not necessarily have a mental disorder, such as family and friends inquiring on behalf of others, or individuals maintaining wellbeing (Fox et al., 2000).

However, individuals with a mental health condition and/or poorer health status are more likely than relatively healthy individuals to access online health information and care (Baker et al., 2003; Powell & Clarke, 2006). Goldner (2006) suggests this is because people who are unwell

are more invested in learning about their illness and associated factors (e.g., treatment and service providers). Furthermore, the open and convenient nature of the web facilitates this while offering a sense of personal control and autonomy that underscores the self-help approach (Christensen & Hickie, 2010) typically adopted by people experiencing mental illness.

Not surprisingly then, surveys of people with a history of mental illness have revealed frequent internet use for mental health information (Burns, Davenport, Durkin, Luscombe, & Hickie, 2010; Khazaal et al., 2008; Lam-Po-Tang & McKay, 2010). A cross-sectional survey of 2000 Australian young adults found that a third of those who had experienced mental health problems in the past five years reported using the internet to help address their problem (Burns et al., 2010). A recent study of 196 psychiatric outpatients in private practice showed almost 95% had internet access (Lam-Po-Tang & McKay, 2010). Over three quarters of this internet using group reported searching mental health-related information, most commonly (all above 75%) regarding symptoms, diagnosis, treatment, and side-effects (Lam-Po-Tang & McKay, 2010).

2.2.2 What draws people to the internet?

The appeal of accessing online mental health resources has been attributed to several factors that underscore the popularity of internet use in general. In the Pew Internet and American Life project survey, the majority of participants seeking health information online reported valuing the convenience, anonymity, and greater range and depth of information afforded by the web (Fox et al., 2000). Indeed, convenience is a frequently reported benefit of using the internet (e.g. Eysenbach & Kohler, 2002; Fox et al., 2000; Umefjord & Hamberg, 2003). In a survey of consumers at a Swedish public health website, “convenience” was endorsed by over half the sample as the main reason for accessing the site (Eysenbach & Kohler, 2002).

The convenience of online resources is apparent when compared to accessing traditional mental health services, such as in-person consultation with a doctor. Rather than experiencing potential difficulties associated with appointments (e.g., lengthy waiting list, travel distance and time), individuals can access the internet at any time, in a range of everyday settings, and at a relatively low cost. With the recent growth of mobile technology (e.g., internet access via mobile phones and tablets), online access has become even more convenient. Another benefit is the ability to obtain, compare, and verify large amounts of information across the web in a timely manner (Eysenbach & Kohler, 2002). When online, consumers can find relevant health information in an average of approximately five minutes (Eysenbach & Kohler, 2002), while over

half of a sample of drug and alcohol website users were reported to take within 5 to 15 minutes to obtain desired information (Klein et al., 2010).

However, online mental health resources are not accessible across circumstances. There can be practical usage barriers such as the lack of affordability (e.g., for low-income individuals) and availability (e.g., in remote regions). Psychological barriers can include negative perceptions and beliefs about computers and the internet (i.e., perceived usefulness and complexity) or limited technical ability (e.g., computing skills) and other personal abilities (e.g., cognitive ability such as attention; hand-eye coordination, and movement) to navigate through the web (Morrell, Mayhorn, & Bennett, 2000). Many of these barriers are more relevant for groups such as the older population (Adams, Stubbs, & Woods, 2005), people with serious mental illness (e.g., schizophrenia; Borzekowski et al., 2009), ethnic minority groups (Daly, 2005), and novice computer users of lower socioeconomic status (Kontos, Bennett, & Viswanath, 2007).

For example, lower rates of online access have been observed amongst individuals aged 55 years and above compared to the general population (Adams et al., 2005). Internet beliefs are considered the main psychological barrier for older individuals, with those avoiding the internet less likely to view it as being useful, easy to use, and efficacious, compared with those who do use it (Adams et al., 2005). Regarding minority groups such as indigenous Australians, the barriers to online mental health resources have been linked to lower levels of education and income associated with these groups (Daly, 2005). As education and income can be closely related, they may have separate or combined effects in explaining internet use (Daly, 2005).

Another online characteristic commonly cited as an incentive of internet use is anonymity. Online mental health resources offer far greater anonymity and privacy than more direct services, such as face-to-face consultation or phone services. The internet can be used without interpersonal contact or directly providing identifying information. While some websites request personal information (e.g., email address for registration), those concerned with privacy can still easily conceal their identity. The relative anonymity of the internet appears to be a particular benefit for online users seeking information about sensitive topics. For instance, 16% of online health seekers reported specifically using the web to research information about topics (e.g., substance use, sexual behaviour, and suicidal ideation) which they found difficult to discuss with others (Fox et al., 2000). Similarly, people with more stigmatised conditions such as mental disorders have been found to more likely use the internet for health information, communicate with clinicians, and seek referral information, compared with less stigmatised health conditions (Berger, Wagner, & Baker, 2005). In recognition of these issues, many mental health-related

websites intentionally minimise barriers of use in terms of compulsory registration details and other user requirements that might compromise convenience, accessibility and anonymity.

2.2.3 Online mental health interventions.

Given the popularity and benefits of the internet, a vast amount of online mental health resources have been made available covering most mental health issues. Amongst interventions available, there is heterogeneity across domains such as whether (Barak, Hen, Boniel-Nissim, & Shapira, 2008): they are automated and website based or utilise online communication (e.g., email counselling); it involves individual or group based therapy (e.g., via a group instant messaging); the website is open to the public or restricted to private access; communication is synchronous (e.g., instant messaging, webcam, audio) or asynchronous (e.g., email, messageboard forum); the modality of content involves text, audio, video/webcam etc; therapy is or is not supplemented with support (e.g., email); and whether therapy follows a particular approach (e.g., CBT, supportive counselling).

Following a comprehensive literature review, Barak, Klein, and Proudfoot (2009) conceptualised internet-supported interventions as consisting of four main categories: 1) online counselling and therapy; 2) internet-operated therapeutic software; 3) web-based interventions; and 3) other online activities. Online counselling and therapy refers to therapy involving interpersonal communication via textual (email or instant messaging) or non-textual means (e.g. videoconferening). Internet-operated therapeutic software represents more advanced programs such as those used for robotic simulation of therapy, rule-based expert systems (i.e., for purposes such as assessment and treatment selection), and gaming or virtual environments (e.g., Second Life). Web-based interventions describe self-guided programs for educational or therapeutic purposes. The fourth category, “other online activities” comprises remaining, various activities that can have therapeutic effects, such as: listening to podcasts; participating in blogs, social networks, and peer-support forums; and gaining knowledge through general websites (e.g., Wikipedia) or online assessment tools (Barak et al., 2009).

Despite recent emergence, internet-supported interventions have received strong support in the literature. A meta-analysis of internet-based interventions for a variety of problems found an overall mean-weighted effect size (.53) similar to the average effect size of traditional, face-to-face therapy (Barak et al., 2008). Other meta-analyses have noted treatment effect sizes for web-based CBT interventions that are relatively smaller for depression-targeted programs (fixed effect size, d = .27), and larger for anxiety-targeted programs (fixed effect size, d = .96; Spek, Cuijpers,

Riper, Keyzer, & Pop, 2007). At present, there is evidence supporting the efficacy of web-based programs for mental health problems such as panic disorder (e.g., Klein, Richards, & Austin, 2001); PTSD (e.g., Klein et al., 2009; Litz, Williams, Wang, Bryant, & Engel, 2004); depression (Christensen, Griffiths, & Jorm, 2004); eating disorders (Ruwaard et al., 2012), alcohol problems (Riper et al., 2011) and other addictions (Gainsbury & Blaszcynski, 2011).

Bolstered by evidence of efficacy, feasibility, and general utility, internet-supported interventions have been promoted within a population-based approach to the prevention and early intervention of mental illness (Christensen & Hickie, 2010). They are also recommended as a low-intensive treatment option in a stepped-care model of mental health service delivery (e.g., Emmerlkamp, 2005; van Straten et al., 2010). The stepped-care approach generally recommends administering the least restrictive treatment and stepping up to a more intensive option if the current treatment is ineffective (Bower, 2005). For example, van Straten et al. (2010) outlined a stepped care model for depression in primary care which recommended guided self-help and internet-based interventions as a follow-up for those with persistent depression symptoms after the initial step of "watchful waiting". If unsuccessful, the model then recommended brief face-to-face therapy, with longer term therapy and/or medication as the final step. This model is associated with large preventive effects in the incidence of depression for older adults displaying sub-threshold depression (van Straten et al., 2010).

2.2.4 Some limitations of internet interventions.

An issue for consumers considering online resources is determining which to utilise. There is wide variability in the content and quality of mental health websites (Ferreira-Lay & Miller, 2008; Griffiths & Christensen, 2005) which places responsibility on individuals to decide on what is suitable for them (Bensley & Lewis, 2002). This process may be especially challenging for people lacking mental health literacy and professional support. Thus concerns have been raised about the impact of low quality information on consumers (Christensen & Griffiths, 2003). For example, inaccurate or inappropriate advice may result in harmful follow-up behaviour and poor treatment choices (Barak, 1999). These issues highlight the need for high-quality web resources as well as a means of connecting people to the appropriate resources.

A consequence of an unsuitable or ineffective intervention can be attrition. Whether referring to non-usage or complete drop out, attrition is commonly observed in studies of web-based interventions and is considered an important challenge to the effectiveness of online interventions (Eysenbach, 2005). In Eysenbach's (2005) seminal review, a number of factors

were proposed to influence attrition including: poor program usability and design; incongruity between user expectations and actual program; ease of drop-out; lack of positive feedback and encouragement; availability of alternative interventions; and external events. Empirical studies of online interventions have also identified predictors of attrition including demographic variables (e.g., employment, marital status); and individual factors such as health locus of control, preferred method of learning, and motivation (Geraghty, Wood, & Hyland, 2010; Klein, Meyer, Austin, & Kyrios, 2011).

With a focus on understanding attrition, Chiu and Eysenbach (2010) conceptualised stages of use which draw on two well-regarded theoretical models: Andersen's (2005) Behavioural Model of Health Service Utilization, which purports that service use is determined by predisposing (e.g. demographics, beliefs), enabling (e.g. community resources), and needs (e.g. perceived needs) factors; and Venkatesh's Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh, Morris, Davis, & Davis, 2003), which posits that the intention to use a program is influenced by constructs such as performance expectancy (e.g. belief that technology will improve performance), effort expectancy (i.e. ease of using new technology), social influence (e.g. recommending of program use), and facilitating conditions (available support to assist use).

Specifically, Chiu and Eysenbach's model emphasises four dynamic and continuous stages of use. The consideration stage refers to the process of deciding whether to use a program, while the initiation stage refers to the period leading up to program commencement. The utilisation stage represents the phase in which the user actively engages with a program, and can result in attrition or continuation (Chiu & Eysenbach, 2010). Finally, the outcome stage refers to outcomes of program interaction, and include clinical outcomes as per Andersen's model, or performance related outcomes as per Venkatesh's model. Chiu and Eysenbach conducted a preliminary study of their model in relation to family caregivers' use of an e-mail support service. They found certain factors more relevant to particular stages. For instance, technological factors (e.g,. perceived ease of use, acceptance of technology) mainly influenced early usage stages, while both clinical factors (e.g., low caregiver competence) and technological factors were more prevalent in later stages (Chiu & Eysenbach, 2010).

While internet interventions are associated with adequate treatment effect sizes, the reality is that many users do not achieve positive outcomes, in part due to attrition. There is limited research directly examining the process of which internet interventions influence behaviour. Ritterband and colleagues (2009) recently proposed a comprehensive model drawing on multi-disciplinary literature, and outlining nine components (e.g., website characteristics, user

characteristics, support, symptom change etc.) that can mediate the impact of internet interventions on behaviour change (Ritterband et al., 2009). Implicit in the model is the premise that an internet intervention program may lead to varying outcomes for different individuals and situations, and that the underlying factors, mechanisms, and processes can be complex. While based on various theoretical and empirical grounds, the model in its entirety has yet to be evaluated (Ritterband et al., 2009).

2.2.5 Models of delivering internet-assisted interventions.

Given that many factors can influence the success of an online intervention, some form of professional support is perhaps necessary to conduct assessment, determine suitability, and offer appropriate follow-up. At one of the first reported examples of computer-based self-help clinics, applicants were screened via telephone interview involving a checklist of areas such as motivation, psychotic symptoms, language ability, and substance use (Gega, Marks, Mataix-Cols, 2004). More recently, a preferred model for delivering web-based mental health intervention programs is with the support of primary care or local GP (Christensen & Hickie, 2010). For example, many mental health websites recommend or require the user to contact a health professional for formal consultation. Alternatively, an individual might see their GP who, after assessing the individual's circumstances, recommends they try an online self-help resource for additional assistance (Buchanan, 1999).

A formal version of this model, as seen at VirtualClinic (https://www.virtualclinic.org.au/), requires clinician prescription as a prerequisite for program access. Inherent in this approach is the implication that the consumer requires an assessment by a health professional before or whilst engaging the online intervention. This is perhaps appropriate in terms of duty of care given the potential issues (e.g. poor quality, ineffective) associated with online interventions. Also, many people, including web users, have an existing relationship with a health professional (Fox et al., 2000; Van Ameringen et al., 2010). However, the reliance on health professional involvement could act as a bottleneck for online service access given the numerous barriers of accessing a health professional (Christensen & Hickie, 2010).

An alternative service delivery model places greater emphasis on open access and self-help. One example is the e-hub web service (www.ehub.anu.edu.au) which provides mental health self-help material through an intervention program and online peer-support forum (Bennett, Reynolds, Christensen, & Griffiths, 2010). Users access e-hub anonymously, free of charge, and without referral or direct assistance from professional staff (though the forum is supervised;

Bennett et al., 2010). This open approach places less emphasis on the need for initial assessment and referral to access the intervention. Instead, it assumes a wide range of people, irrespective of their symptom levels and whether they have received a formal diagnosis, will still benefit from the intervention.

A variation of this self-help model involves access to more formal and intensive services, such as in-depth clinical diagnostic assessment and CBT-based treatment, without necessary involvement of external health professionals. The terms “e-mental health clinics” or “virtual e-clinics” are associated with this approach (Christensen & Hickie, 2010). One example is Anxiety Online (AO; described further in Chapter Three) which offers both online assessment and self-help or therapist-assisted treatment programs, free of charge to the public (Klein et al., 2011). While many online services avoid providing diagnostic results and assigns this to health professionals, AO integrates the diagnostic process into its service delivery model. An individual reluctant or unable to see a health professional, can therefore undertake a web-based diagnostic assessment, receive referral suggestions, and go onto complete an intervention program within the same website. While there is evidence supporting this approach (Klein et al., 2011), further research is needed to confirm its overall effectiveness.

2.3 Online clinical assessment.

The growth of the internet and heterogeneity of internet-assisted technologies have inspired a range of online activities related to clinical assessment, with the general purposes of aiding clinicians and researchers in conducting clinical assessment, and promoting mental health awareness amongst self-help consumers (Barak & Buchanan, 2004). These activities vary from simple web-based questionnaires for screening mental disorders, through to internet-assisted videoconferencing for diagnostic clinical interviewing. The internet has provided several characteristics generally not possible with purely computerised assessment. For instance, the internet allows for physical distance between assessors and consumers throughout the assessment process (Bartram, 1997; Sale, 2006).

Furthermore, the internet allows the centralised storage of assessment tools and decentralised access from a remote location (Naglieri et al., 2004). Through the automation of website applications, the internet has also allowed the removal of the clinician from the assessment process, which contrasts to other indirect forms of assessment (e.g., palm top computers, mailed paper-pencil questionnaire, telephone interview) that still require some clinician involvement. The reduced reliance on supervision has been argued to have both a

positive impact (e.g., enhancing self-disclosure) and negative impact (e.g., lack of control over invalid responses) on online assessment and these issues will be discussed later in this chapter.

In the following sections, common examples of online clinical assessment will be introduced under two categories: internet-assisted clinical interviewing; and web-based clinical assessment. The former category essentially involves a clinician conducting assessment via internet-assisted communication, while the second category predominantly relies on automated web-based programs to administer questionnaires and other testing instrument. Here, web-based assessment describes any program delivered through the web which is largely automated and can be undertaken without direct involvement from a clinician or interviewer. It is acknowledged there are other online resources and activities that could aid clinical assessment (e.g., see Barak & Buchanan, 2004). However, a review of these would be too exhaustive to consider here, though some exciting opportunities will be mentioned in Chapter Seven.

2.3.1 Internet-assisted clinical interviewing.

The internet has been used in several ways to facilitate clinical interviewing between interviewer and interviewee. One aspect of interaction is whether it is *asynchronous* or *synchronous*. Asynchronous interaction refers to the exchange of information at different time points (e.g., email exchange). While asynchronous forms are likely used by many professionals (particularly conducting email counselling) there are no known studies of their effectiveness for the specific purpose of clinical assessment. Nevertheless, the ability to respond at different times could enhance convenience and promote more thoughtful, precise, and relevant communication, as there is time to carefully consider questions and construct responses (Tate & Zabinski, 2004). However, such communication seems unsuitable for an immediate measure and response to a current situation, and is likely more conducive for longitudinal assessment of mental health.

With synchronous methods of online communication, the interviewer and interviewee are interacting at the same time (“real time”) in a dynamic exchange. Synchronous clinical interviewing could involve text based formats such as instant messaging or online “chat”, voice over Internet Protocol (VOIP), and videoconferencing (Barak & Buchanan, 2004). The general benefits and disadvantages of instant messaging for interviewing were recently reviewed in a study examining its use in qualitative research in psychology. Specifically, Jowett, Peel, and Shaw (2011) discussed the process of using instant messaging for interviewing homosexual and bisexual individuals living with diabetes. The authors highlighted how instant messaging afforded convenience in accessibility, but resulted in longer interview times despite less data quantity than

elicited in a face-to-face interview. Issues in establishing rapport, listening, expressing emotions, and technical requirements in instant messaging where also identified, and resemble those commonly associated with online counselling via instant messaging (e.g., Bambling, King, Reid, & Wegner, 2008; Williams, Bambling, King., & Abbott, 2009).

In a rare study demonstrating the use of instant messaging for clinical diagnosic assessment, Mallen, Jenkins, Vogel, and Day (2011) examined the performance of online counsellors in being able to diagnose DSM-IV disorders from online sessions (45-50 minutes) with clients. Counsellors were graduate students in advanced training while confederates performed the role of individuals presenting particular symptom profiles. The study found counsellors correctly identified 90% of depression cases and 86% of anxiety cases, though accuracy dropped for specific diagnoses (e.g. social phobia) and mixed symptom presentations (Mallen et al., 2011). Mallen et al. suggested online chat could be used to accurately assign a DSM-IV diagnosis of straightforward presentations; however, mixed presentations may be misdiagnosed due to absent non-verbal cues and the inherent difficulty of diagnosing complex cases. Bambling et al. (2008), who qualitatively studied experiences of online counsellors, suggested assessment is more difficult through instant messaging compared to face-to-face due to the reduced emotional promixity between the client and the assessor; in other words, clients reveal less as they feel less connected with their counsellor.

Of the different online modes, “telepsychiatry”, which refers to the use of videoconferencing for psychiatric interviewing (García-Lizana & Muñoz-Mayorga, 2010), has been widely investigated and offers the closest relation to traditional clinical interviewing conducted in person. This is because it maintains visual and audio communication despite geographical distance between clinician and client. In doing so, important processes involved in face-to-face interviewing are preserved, such as assessing visual and auditory cues presented by the client, and the building of rapport through body language. Hyler et al. (2005) conducted a meta-analysis of 14 studies comparing the diagnostic results of telepsychiatry and in-person psychiatric assessment. Agreement between the intraclass correlation coefficients of in-person assessment and telepsychiatry were found to be excellent (Hyler et al., 2005). A correlation of inter-rater agreement between in-person and telepsychiatry was also considered very high. There were no significant differences in satisfaction measures, though patients and clinicians stated a stronger preference for in-person assessment (Hyler et al., 2005).

Given these results, telepsychiatry has been suggested as a means of addressing the shortage of specialist healthcare professionals in rural areas (Hyler et al., 2005). However,

telepsychiatry has notable limitations which prevent its widespread use. Firstly, it has significant technological requirements (e.g., hardware, bandwidth in two locations) which could be costly for consumers, though recent proliferation of internet-based video calling (e.g., via Skype, FaceTime) could improve access. The reliability of telepsychiatric diagnosis can also impaired by lower quality internet bandwidth (Yoshino et al., 2001), though further evidence is needed to confirm this (Grady, Myers, & Nelson, 2009). Telepsychiatry, as with other synchronous methods, also still requires an interviewer to conduct the assessment, which is arguably the most costly expense of a traditional clinical assessment. There is suggestion that telepsychiatry could also promote excessive demand for certain professionals (e.g., more well regarded) at the neglect of others, perhaps more readily available (Hyler et al., 2005). Consequently there are doubts about telepsychiatry's cost-effectiveness (i.e., due to high ongoing costs; García-Lizana & Muñoz-Mayorga, 2010) and the pressure it places on a limited healthcare professional labour force (Hyler et al., 2005).

2.3.2 Web-based clinical assessment

This category refers to largely automated programs, often based on existing paper-pencil measures, delivered via the web, and intended to assist the clinical assessment process. In the tradition of paper-pencil questionnaires and previous computerised interviews, they generally comprise of text-based multiple-choice items (Buchanan, 2002), though some programs have incorporated audio and visual components (e.g., Chinman et al., 2007; van Ballegooijen et al., 2012) or relied on typed out responses which are automatically processed (Steiger & Reips, 2008). Access of such programs varies in the level of external control and supervision. However, most applications reported in the literature or found online appear to follow an *open* (i.e., completed with no supervision or registration) or *controlled* mode (i.e., requires some registration but completed without supervision) as defined by the ITC (2005).

Web-based clinical assessment programs have been proposed for use in several clinical/counselling related scenarios consistent with the main purposes of clinical assessment: to assist decision-making; for screening and diagnostic purposes; for the monitoring of symptoms, and for the planning and evaluation of treatment (Buchanan, 2002; ITC, 2005). Specifically, web-based clinical questionnaires have been highly utilised in research contexts (e.g., in place of paper-pencil symptoms scales) and also used by consumers to facilitate self-exploration and self-awareness in the context of self-help (Barak & Buchanan, 2004). More recently, researchers and developers have begun to couple web-based assessment with online treatment programs to form

wholly online clinics, therefore allowing clients to bypass the barriers associated with traditional clinician-administered services (Klein et al., 2011).

In previous literature, the programs considered as part of web-based assessment have also been referred to terms such as “online testing”, “internet-based testing”, “online questionnaires”, “web-based screening”, and “web-based assessment” depending on the assessment context (Barak & Buchanan, 2004; Buchanan, 2002; ITC, 2005; Maheu & Gordon, 2000). From here on, these terms will be used interchangeably with respect to how they are presented in associated literature; however, they will all be considered as part of the broad category of web-based clinical assessment (Buchanan, 2002), and in short will be referred to as “web assessment”. As the focus of the present research involves a web assessment program, the remainder of this review will concentrate on literature relating to this area. Of the various applications of online clinical assessment, web-based programs have arguably been the most actively pursued by consumers, clinicians, and researchers due to their potential benefits.

2.3.3 The benefits of web assessment.

Various factors are regularly cited to explain the appeal of delivering health interventions via the internet which extend to web assessment (Griffiths, Lindenmeyer, Powell, Lowe, & Thorogood, 2006). As with computerised assessment, many benefits are due to the programmable nature of an online program, which allows built-in item branching, immediate scoring processes, and automated provision of feedback (Barak & English, 2002). However, improvements in technology now allows websites to have far greater functionality than earlier computer programs, such as integrated multimedia (e.g., text, sound, video, animation) and interactivity (e.g., navigating through webpages, inputting text/audio/video, interacting with diagrams, controlling virtual characters).

Current online programs also have capacity to be tailored to the individual and, for example, adjust to the preferences of the user and provide more meaningful content (Griffiths et al., 2006; Noyes & Garland, 2008). For researchers, the computerisation of data collection minimises data-entry errors and the issue of missing values, (i.e. by programming items to require a response before submission; Austin, Carlbring, Richards, & Andersson, 2006). Web assessment programs also have the potential to covertly measure usage data (e.g., frequency of use, content preferences) and cognitive/behavioural data (e.g., time to perform a task; decision making processes; Noyes & Garland, 2008).

There are also unique benefits facilitated by the internet that were not generally associated with earlier computerised assessment. Firstly, due to widespread internet use, an online program has the potential to reach a large number of people (Buchanan, 2000, Sale, 2006). This allows for more diverse samples when conducting research (Schmidt, 1997), while for clinicians, it allows service delivery to those who are unable (e.g., due to disability), or refuse to attend traditional services such as an in-person clinic consultation (Buchanan, 2002). Due to the automated nature and ongoing availability of the web, online programs can also continuously operate, therefore catering for people in different time zones or with alternative lifestyles.

Another appealing characteristic of web assessment is that program content and respondent data can be centrally stored and managed, despite decentralised use by consumers. Accordingly, it is far easier to update or revise an online assessment program (e.g., in response to psychometric evaluation) or extract and process large amounts of data (e.g., for research purposes) compared to dealing with physical means such as hard copies of printed interview and questionnaire forms (Naglieri et al., 2004). Last but not least, the online environment can provide a greater sense of anonymity which elicits disinhibition and facilitates self-disclosure (Joinson, 2001). This interaction may play a special role in the accuracy of online assessment and will be discussed in greater detail later in this chapter.

The financial benefit of online assessment.

Although initial development of an online program can be expensive, running and maintenance costs are relatively low compared with other resources such as printed questionnaires and clinician-involved services at a physical clinic. The term “scalable”, which describes the capacity to increase service volume with relatively little additional cost, is often associated with the internet and aptly applies to the cost-efficiency of web-based assessment (Naglieri et al., 2004). To demonstrate this, the number of people completing an online questionnaire could, in one year, increase by 5,000 administrations but create little extra cost to the developer except maybe additional server space. In contrast, paper-pencil questionnaire administration in a clinic would require an additional 5,000 questionnaire forms, and extra clinic staffing and resources to administer, score, and store questionnaires.

Houston et al. (2001) developed an online version of the CES-D for public screening of depression. Over eight months, the program was completed 24,479 times. Given the sunk (i.e., development) cost of $9,000 and marginal (i.e., maintenance) cost of $468 per month, the online CES-D program cost an estimated $0.52 per administration which was considered cost-effective

by the authors. Griffis, Goldsby, and Cooper (2003) compared the cost of administering a survey by mail (e.g., costs of printing, postage), and via the web (e.g. webpage and database development), while factoring the number of received responses. The results showed the cost per usable response was $38.64 for the mail survey (*n* = 1538) and $23.62 for the online survey (*n* = 2411). The expenses did not include data entry costs, which the authors suggested would improve the cost advantage of the online survey. It was concluded that online surveys offered a more cost-effective means of administration than posted questionnaires, and the advantage was due to the higher online response rate which reduced marginal costs as respondent numbers increased (Griffis et al., 2003).

Thus, if the purpose of an assessment tool is to reach a large number of people at a low cost, the online medium offers an advantage over the postage means. There are no known studies comparing the costs of web-based assessment against interviewer-administered assessment. However, Carlbring et al. (2007) suggest that for research studies involving a clinical screening process for potential participants, replacing telephone and in-person interviews with equivalent online assessment measures could save time and money given that many individuals are usually ruled out after the screening process.

2.3.4 The disadvantages of web assessment for clinicians and researchers.

Disadvantages of administering web-based assessment have also been clearly noted (Barak & English, 2002; Barak & Hen, 2008; Buchanan, 2002; Donker, van Straten, Marks, & Cuijpers, 2009; Naglieri et al., 2004). Further to the drawbacks of computerised assessment, a major drawback of online assessment is that it can lack control over testing conditions. For example, when respondents are anonymous and/or unsupervised (e.g., accessing an online questionnaire from home), they are at risk of compromising data quality by feigning their responses, responding frivolously or maliciously (Donker et al., 2009), or participating in an inappropriate state (e.g. distracted or intoxicated). Without supervising respondents, assessors/administrators cannot monitor these issues and factor them into assessment results, nor promptly respond to participants experiencing adverse effects (Reips, 2002). Clinician absence also denies important components of clinical interviewing such as assessing mental state (e.g., behaviour, mood, non-verbal emotional expression, cognition), applying clinical intuition and judgment to subtle but significant cues, and fostering a therapeutic relationship (e.g., rapport) between client and practitioner (Taylor & Luce, 2003).,

Web assessment programs can also pose accessibility difficulties. As mentioned earlier, many people do not have readily available internet to access online assessment options. As for computerised assessment, poor computer literacy and/or discomfort towards computers and the internet could also limit access and compromise performance for certain people (Noyes & Garland, 2008). Furthermore, technical factors can create variability in the user experience. Challenges on the server-side include ensuring that a website is stable, has sufficient capacity for expected demand, and can run across different internet and computer platforms (Reips, 2002). User-side problems include insufficient or incompatible hardware and software or having malicious software (e.g. virus, malware) which interferes with the online program. Moreover, internet activities are vulnerable to security breaches (e.g., website hacking) which can expose data and compromise privacy (Naglieri et al., 2004). For developers, there are also issues related to copyright (Barak & English, 2002), including the difficulties of acquiring licensing to use existing measures, and protecting the copyright of one's own program especially when it is openly available to the public. This point in particular has been suggested to account for the limited offering of well-established clinical assessment instruments via the web (Butcher, 2006).

For text-based web assessment, there is also added risk of misinterpretation. Poorly worded content (e.g., questionnaire items) and a lack of supervision to clarify the meaning of items could result in misinterpretation and misrepresentation by respondents, particularly from culturally and linguistically diverse groups (Barak & English, 2002). Inaccurate or overly generalised assessment feedback could also be misconstrued. Currently, many online clinical diagnostic tests use a "typology" approach which automatically generates standardised reports corresponding to a respondent's results (Butcher, 2004). This draws on the programmable nature of the web, but is limited by the sophistication of the programming. As Sale (2006) raises, this approach can reduce the nuance and person-specific nature of test results across different people with highly varied attributes and situations. Hence, through over-generalisation, a web assessment program could produce results that misrepresent the individual.

2.3.5 Ethical issues of web assessment

Several reviews have highlighted a range of professional and ethical issues associated with online testing that could relate to some of the previously reviewed web assessment programs (Barak & English, 2002; Barak & Buchanan, 2004; Buchanan, 2002; Reip, 2002; Naglieri et al., 2004). For instance, there is often an inadequate professional relationship between consumers and the developer/administrator of web programs influenced by the physical and

psychological distance afforded by the internet (Barak & English, 2002). Consequently, it seems many developers show little accountability and responsibility towards consumers (e.g., in terms of offering adequate support throughout testing), while consumers are at risk of abusing a program and its results (e.g., responding in a careless manner or repeating a test to obtain different results; Barak & English, 2002).

Given the ease of offering web assessment, there are also concerns about unqualified people developing assessment procedures without sound theoretical or scientific principles, and therefore facilitating the dissemination of assessment results of questionable accuracy and reliability (Barak & English, 2002; Buchanan, 2002). As many programs are not accompanied with guidelines for use and evaluative evidence (e.g., of psychometric properties), they are susceptible to being used and interpreted out of context by unqualified and unsuspecting users. For example, one could imagine an emotionally vulnerable individual undertaking an online clinical questionnaire, receiving a disturbing result, and experiencing further distress (Buchanan, 2002). Ethical issues such as these could be pronounced in relation to online programs offered in an open and uncontrolled manner which, at present, represent many of the top search web-search results as well online instruments made available by publishers and developers (Sale, 2006).

2.3.6 Online testing guidelines.

In response to the growth of web assessment and professional and ethical issues associated with it, various guidelines have been proposed for the use of online clinical assessment instruments (ITC, 2005; Naglieri et al., 2004; Reips, 2002; Suler, 2001), often reflecting previous professional and ethical guidelines applied to standard paper-pencil psychological testing (e.g., APS, 2009). In particular, the International Testing Commission (ITC; 2005) published extensive guidelines to raise awareness amongst stakeholders of the issues of good practice in computer and internet based testing (Sale, 2006). Directed to developers, publishers, and test users (e.g. clinicians, consumers), the ITC guidelines outline four key areas to consider in employing computerised and online testing (ITC, 2005): ensuring technology needs are considered and met; ensuring the quality of testing material and processes; controlling the administration of testing; and maintaining the security of testing materials and data.

In general, the ITC (2005) emphasise appropriate use of online testing, and that users are informed with a clear statement of the limitations of a specific online test. As part of this, the guidelines recommend the documentation and dissemination of psychometric information, and that only online tests with evidence of adequate psychometrics are offered (ITC, 2005). It is also

advised that care be taken to ensure an online instrument "does not require knowledge, skills, or abilities…that are irrelevant to or might impede the test-taker's ability to perform the test". In other words, there is an emphasis on ensuring the match between the needs and abilities of respondents and the characteristics required for an online test. To inform these issues, the ITC guidelines advise that evaluation "be conducted over the Internet with participants completing the test under conditions that represent those that the intended target population will experience (e.g., unstandardised testing conditions)" (pg. 11). In doing so, stakeholders can obtain an accurate reflection of the program's actual performance.

2.3.7 Current web-based clinical questionnaires.

Web assessment programs emerged soon after the web launched in the mid 1990s. In their review of online assessment, Barak and Hen (2008) described how early online psychological tests were basic in content and seemingly created by non-professionals. However, the growing acceptance of the internet and recognition of its advantages has lead many consumers, professionals, and organisations to offer a range of publically available web-based clinical instruments of varying attributes and quality. To exemplify this, an Australian Google search using the search terms "depression test" resulted in approximately 130,000,000 search results (www.google.com, accessed 24/4/2012); though many are likely irrelevant or in reference to the same webpage or online program.

The top result linked to the *beyondblue* website (www.beyondblue.org.au, accessed 24/4/2012) which promotes depression and anxiety awareness. The website offered three "checklists" for depression: a generic nine-item measure of depression symptoms; the well known Kessler Psychological Distress Scale (K-10) comprising 10 items with a 5-point likert scale (Kessler & Mroczek, 1994); and the SPHERE questionnaire, consisting of 27 items covering depression symptoms across different domains (Clarke & McKenzie, 2003). Completing each checklist resulted in the presentation of: a score based on the number of symptoms/items endorsed; a brief guide to clinical threshold (e.g., for the nine-item checklist: "If you scored five or more you may have depression"), a statement that checklists should be used as a "rough guide"; and a recommendation to see a doctor for a full diagnosis. While overall content was brief, a reference for each checklist was included that directed to relevant research..

The remaining top search results included other questionnaires that varied in length, complexity, and credibility. Several comprised of 10 dichotomous items corresponding to DSM-IV-TR criteria of a major depressive episode, while few programs delved into functional impairment,

suicidality, medication use, and the potential influence of bereavement, medical illness, and substance use on depression symptoms. The lengthiest test amongst the top search results was offered by Psychology Today (http://psychologytoday.tests.psychtests.com/take_test.php?idRegTest=1308, accessed 24/4/2012). This consisted of 101 items (mostly rated on a 5 point scale), covering different facets of depression and general functioning, and was suggested to take 30 minutes. The test ended with a general description of depression, and a "snapshot" of one of its measures (e.g. "worry about judgement of others") consisting of a score and brief interpretation. However, the site requested a payment of USD$4.95 for a full report.

When a search was repeated for "anxiety test", more diverse measures appeared in the top results, including: the "Albert and Haver's Achievement Anxiety Test"; the Sport Competition Anxiety Test at www.brianmac.co.uk/scat.htm; and the Liebowitz Social Anxiety Scale Test at www.socialanxietysupport.com/disorder/liebowitz/. Several websites in the top results for "depression test" reappeared with anxiety related questionnaires. For example, at the beyondblue website, the three previously mentioned checklists for depression were also recommended as anxiety measures. In addition, the main anxiety disorders each had checklists for users to endorse relevant symptoms. Scores were not provided, and instead, the site highlighted the importance of seeing a doctor if experiencing any symptoms (Beyondblue, 2012)

The top result for an "anxiety test" referred to Queendom (http://www.queendom.com/tests/, accessed 24/4/2012), a website promoted as "The land of tests". The Queendom anxiety test had 42 questions referring to various anxiety symptoms rated on a 5 point scale, followed by sociodemographic questions (e.g. age, ethnicity, education). The test also provided a brief report of an anxiety subscale (e.g. "existential anxiety"), and charged USD$6.95 for a full report containing five subscale scores and interpretations. A "scientifically validated" stamp was listed, which linked to descriptive statistics and percentile scores of a sample with 7731 subjects. However, the vague document did not explain terms, details about the sample, and other important psychometric information. Nevertheless, as previously noted by Barak and English (2002), the Queendom website is one of the few websites in the top searches to directly link to psychometric results.

In summary, most of the web assessment tools found through the brief web search consisted of a questionnaire format using predominantly closed-response items which resemble paper-pencil questionnaires and computerised assessment programs that precede it (Buchanan, 2002). Most websites included some brief disclaimer about the limitations of the test. A few required the user to tick a box acknowledging their use of the test for informal reasons (e.g.,

"entertainment purposes", "personal use") and that proper diagnosis should be made through a health professional. Furthermore, programs typically provided limited feedback in terms of a score and general advice about consulting a health professional (Buchanan, 2003). Almost all did not offer referral information (besides contacting a doctor) or recommendations of further online resources to consider. The majority were free to access, while lengthier and more elaborate looking tests often required a small fee to obtain an assessment report.

Although Google search results may not reflect the depth in quality and different types of online clinical measures available, they can indicate which web-based programs are popular according to search engine algorithms (Page, Brin, Motwani, & Winograd, 1999). People with little internet experience are unaware of how search engines operate (Fallow, 2005). Many consumers rely on search engine recommendations, and few likely consider alternatives beyond top results. This suggests the online programs aforementioned are being offered to and utilised by numerous people seeking online measures of anxiety and depression. It is unclear as to how beneficial many of these programs are for users, particularly when there is little evidence of evaluation. Fortunately, a number of studies in the mental health literature have examined the psychometric properties of certain web-based clinical assessment measures and other aspects of use.

2.3.8 Psychometric considerations of online clinical assessment.

There are several methods of measuring the reliability and validity of an online clinical assessment program, each varying in suitability depending on the type of program involved. Web-based questionnaires that replicate the content of paper-pencil questionnaires can employ the same psychometric measures for reliability (e.g., internal consistency, test-retest reliability) and validity (e.g., factor analysis, criterion validity). But for programs that tailor to the individual by involving branching/skipping rules dependent on previous responses, many standard reliability measures are generally difficult to assess. For example internal consistency, which represents the uniformality of items within a test and is typically measured by Cronbach co-efficient alpha or split-half, is considered inappropriate because item response sets vary among respondents (Ruland et al., 2007).

On the other hand, test-retest reliability can be used for most types of online programs to measure the stability of results (e.g. test scores, diagnostic outcomes) over two separate administrations to the same participants (Ayearst & Bagby, 2010). This can be computed via the Pearson correlation coefficient for continuous data, the Spearman correlation for ranked data, and Cohen's kappa coefficient for categorical data (Ayearst & Bagby, 2010). If the concept being

measured by an instrument is sensitive to change over time, then test-retest reliability is not an appropriate measure of reliability (Ayearst & Bagby, 2010). In the case of tailored programs (i.e., with branching rules), it is worth noting when a minor change in symptom reporting could lead to majors changes in the subsequent item's presented and the eventual diagnostic outcome, as it may greatly influence test-retest reliability (Ruland et al., 2007).

Of the main types of validity (e.g., content, criterion, construct), the criterion validity of an online diagnostic program is often of most interest. Criterion validity is measured by comparing an online program's results with the performance of a criterion measure (Groth-Marnat, 2003). Structured clinical interviews conducted by a clinician are often employed as the "gold standard" criterion of validity, given they display optimal psychometric properties and are reasonably feasible. However, standardised and validated self-report measures have also been considered as validity criteria (e.g. Donker et al., 2008). When assessing the validity of web-administered questionnaires based on existing paper-pencil measures, the original questionnaire is often used as the criterion. In this context, procedural validity reflects the extent of agreement between the results of the original (e.g. paper-pencil) and online administration of an assessment instrument.

Construct validity is also commonly examined and refers to how well a measure performs in accordance with the concepts or constructs being measured. Evaluating construct validity can involve factor analysis, convergent validation, and discriminant validation, amongst other methods. For assessment programs with variable item sets (i.e., due to branching rules), factor analysis is difficult to perform, particular if there is a small sample size (Ruland et al., 2007). However, convergent and divergent validity are considered more appropriate with tailored assessment programs. Convergent validity reflects the agreement between two independent methods that purport to measure the same or similar construct. Discriminant validity refers to the extent of which a test's scores do not correlate with the scores of a test that measure an unrelated construct (Ayearst & Bagby, 2010).

An online program demonstrates high convergent validity if its results associate highly with another measure. Several factors can contribute to high convergent validity. Two instruments might share a large amount of variance and therefore strongly correlate because they measure the same trait, employ the same items, or share the same test methods (e.g., both involving self-report or paper-pencil format). As high correlation between two instruments can be due to either shared trait, shared items, or shared methods, convergent validity needs to be carefully investigated by researchers (Ayearst & Bagby, 2010). There are a range of statistical indices commonly used to measure convergent and discriminant validity which overlap with criterion

validity. For example, indices such as sensitively, specificity, positive and negative predictive power are generally used (Hsu, 2002) and can be classified as part of criterion or construct validity. In general, the various types of validity can fall under the umbrella of construct validity as they provide evidence of the construct being measured by the online program evaluated.

2.3.9 Psychometrics of web-based clinical questionnaires.

The literature has reported numerous evaluations of web-based assessment tools developed for a range of health related topics, including the monitoring of cognitive status (Erlanger et al., 2002), the measurement of cigarette withdrawal (Etter, 2005), traumatic stress (Fortson, Scotti, Del Ben, & Chen, 2006), the impact of headaches (Bayless et al., 2003), and alcohol consumption (Cunningham, Humphreys, Kypri, & van Mierlo, 2006). Amongst these are web-based questionnaires for the screening of mental disorders based on existing paper-pencil questionnaires. These include established measures of depression (e.g., Donker et al., 2010; Holländare, Andersson, & Engstrom, 2010; Spek, Nyklicek, Cuijpers, & Pop, 2008), panic disorder and agoraphobia (e.g., Austin et al., 2006), and OCD (e.g., Coles, Cook, & Blake, 2007). Instead of being created with original content, these programs have been based on existing measures because they are standardised, have associated literature (e.g, regarding psychometrics), and are accepted in clinical and research settings (Donker et al., 2008). Consequently, these web-administered versions resemble the appearance of the original measure, with the aim of eliciting similar results and psychometric properties.

However, several factors have been noted to potentially diminish the psychometric "equivalence" of web-administered questionnaires and their paper-pencil counterpart (Buchanan & Smith, 1999; Buchanan, 2002). Ferrando and Lorenzo-Seva (2005) summarise these factors as relating to "the nature of the sample, the medium of presentation, and the circumstances under which a test is taken" (pg. 193). Firstly, the samples of which psychometric evidence for traditional assessment tools are based on may differ from the population undertaking online versions (Barak & Buchanan, 2004). The reach of the internet allows greater heterogeneity in respondent background (e.g., age, education, language) which could add confounding variables to the outcome of online measures (Buchanan & Smith, 1999), and limit the relevance of existing psychometric evidence (e.g., of norms) derived from paper-pencil questionnaires and using more narrow population samples. For example, McCue, Buchanan, and Martin (2006) examined the Hospital Anxiety and Depression Scale (HADS) and found higher measures of anxiety and depression in their internet recruited sample compared with normative control data.

On the other hand, identical samples could still reach different results on a web-based and traditionally administered assessment tool due to an interaction between individual factors, the computerised medium, and the circumstances under which a test is taken. As earlier described, computer use has been associated with a number of effects in relation to computerised assessment. This include increased awareness of the computer medium, enhanced reading of text, and the potential experience of computer anxiety which could interfere with computer interaction and possibly inflate scores on computerised anxiety measures (Webster & Compeau, 1996; George, Lankford, & Wilson, 1992).

The ability to perform online assessment in a range of open and uncontrolled settings may also create variation due to extraneous factors (e.g., distraction, environmental cues, technical issues and variation in hardware/software) as well as temporary factors (e.g., intoxication), leading to results which differ from those otherwise elicited by completing a questionnaire or interview in more controlled environment's such as in a clinic or research laboratory (Austin et al., 2006; Buchanan & Smith, 1999; Ferrando & Lorenxo-Seva, 2005). With the increased anonymity and relative privacy of typical internet use, the influence of social desirability could also be less pronounced in web administration compared with paper-pencil and in-person administration (Joinson, 1999). For example, an individual completing an online survey from home would presumably feel less inhibited and perform with less social desirability and extrinsic motivation compared with undertaking the same questionnaire under supervision. A consequence of this could be relatively higher scores on web-based instruments collecting information that people are more comfortable disclosing online (Buchanan, 2003).

2.3.10 Studies of equivalence.

Given the possible disparity in the experience and outcome of web-based and paper-pencil questionnaires, the psychometric equivalence between the two forms cannot be assumed (Buchanan, 2002, 2003). Consequently, the literature has examined the psychometric equivalence of many web-based clinical questionnaires and their paper-pencil counterparts usually by administering both forms to the same sample and comparing the results (see Table 1 below). While many studies have shown largely equivalent psychometric properties (e.g. Carlbring et al., 2007; Herrero & Meneses, 2006; Houston et al., 2001; Spek et al., 2008), others have indicated subtle differences in terms of norms and scoring distributions (e.g. Andersson, Kaldo-Sandstrom, Strom, & Stromgren, 2003; Austin et al, 2006; Carlbring et al., 2007; Joinson,

1999; McCue et al., 2006; Vallejo, Jordan, Diaz, Comeche, & Ortega, 2007) as well as factor loadings and structure (Buchanan, 2002; McCue et al., 2006; Riva, Terruzi, & Anolli, 2007).

Table 1

*Summary of Identified Studies Evaluating the Psychometric Properties of Web-Based Clinical Assessment Measures**

Online administration	Author (year)	Design and Methodology	Main psychometric results	Findings
Perceived Stress Scale	Herrero & Meneses (2006)	Equivalence with p-p Randomised order *N* = 530 university students, frequent internet users	No differences in internal consistency No differences in factor structure based on non-significant chi square test	Equivalence with p-p format for university population
Body Sensations Questionnaire (BSQ)	Austin et al. (2006)	Equivalence with p-p Randomised order *N* = 110 people with panic disorder	Strong to very strong correlation between p-p and web Higher mean score for p-p than web, medium-sized effect Similar internal consistency	Mostly equivalent with p-p for panic disorder population
Agoraphobic Cognitions Questionnaire (ACQ)	Austin et al. (2006)	Equivalence with p-p Randomised order *N* = 110 people with panic disorder	Large correlation between web and p-p (*r* = .89) Lower mean score for web than p-p, medium effect size Negligible difference in internal consistency	Mostly equivalent with p-p for panic disorder population
	Carlbring et al. (2007)	Equivalence with p-p Randomised order *N* = 494 applicants of online panic dis. intervention program	Large correlation between web and p-p (*r* = .89) Lower mean score for web than p-p, medium effect size Negligible difference in internal consistency	
Beck Anxiety Inventory	Carlbring et al. (2007)	Equivalence with p-p Randomised order *N* = 494 applicants of online panic dis. intervention program	High correlation between scores (*r* = .84) Lower mean score for web than p-p, large effect size Negligible difference in internal consistency	Mostly equivalent with p-p except in mean scores
Dissociative Experiences Scale (DES)	Collins & Jones (2004)	Equivalence with p-p Between subjects design, *n* = 42 completed p-p version *n* = 293 completed web version	Significantly higher mean scores for web than p-p Similar internal consistency	Evidence of equivalence
GAD-7	Donker et al. (2011)	Criterion validity (CIDI) *N* = 157 adults recruited via internet Focused on range of anxiety disorders and depression	Significantly higher GAD-7 scores for those with CIDI diagnosis of GAD, compared with those without a diagnosis. Good predictive accuracy for GAD (AUC = .77)	Good diagnostic validity for GAD in internet population
Hospital Anxiety and Depression Scale –	Andersson, Kaldo-Sandstrom, Strom, &	Equivalence with p-p *n* = 157 patients with tinnitus completed web version	Higher percentage of clinical anxiety in web (25%) than p-p (15%) Similar internal consistency between web and p-p	Limited evidence - adequate for screening

Anxiety (HADS-A)	Stromgren (2003)	*n* = 86 non-patients completed p-p version		when cut-off of 9 or greater is used in primary care setting.
	Bunevicius, Peceliuniene, Mickuviene, & Bunevicius (2007)	Criterion validity – MINI (anxiety disorders) *N* = 503 primary care patients	HADS-A cutoff score of 9 or greater: Good predictive accuracy for GAD diagnosis (AUC = .79-.81) Sensitivity = .76-1.00 Specificity = .63-.77	
	McCue et al. (2006)	Equivalence with p-p Between subjects design *n* = 494 people with symptoms of Chronic Fatigue Syndrome (CFS), *n* = 1362 without CFS	Higher mean scores for web compared with normative control data collected with p-p version Similar factor structure to previous reports of p-p	
Obsessive-Compulsive Inventory (OBI)	Coles et al. (2007)	Equivalence with p-p *N* = 106 undergraduate students	High correlation between web and p-p (*r* = .89) No significant difference in mean scores between web and p-p No significant order effects	Equivalent with p-p for undergraduate students
Obsessive Beliefs Questionnaire (OBQ)	Coles et al. (2007)	Equivalence with p-p *N* = 106 undergraduate students	High correlation between web and p-p (*r* = 0.83) No significant difference in mean scores between web and p-p No significant order effects	Equivalent with p-p for undergraduate students
Trauma Symptom Screen	Fortson et al. (2006)	Equivalence with p-p Randomised order *N* = 411 university students	Similar internal consistency and means scores between web and p-p, across different time points; differences were not statistically tested	Limited evidence of equivalence with p-p
Beck Depression Inventory-second edition (BDI-II)	Holländare et al. (2010)	Equivalence with p-p *N* = 87 participants from primary and psychiatric care in Sweden	High correlation between scores (*r* = .89) Suicidality item had significantly lower score on internet administration	High correlation with p-p format but potentially different mean scores
	Carlbring et al. (2007)	Equivalence with p-p *N* = 494 applicants of online panic dis. intervention program	High correlation between scores (*r* = .94) Higher mean scores on internet format, small effect size	
CES-D	Cuijpers, Boluijt, & Van Straten (2008)	Criterion validity (MINI) and reliability *N* = 243 adolescents	High internal consistency (.93) High correlation with web MDI (*r* = .88) Strong predictive accuracy of MDD (AUC = .90)	Good validity and internal reliability, suitable use for screening across different population groups
	Donker et al. (2010)	Criterion validity (CIDI – any depressive disorder) *N* = 157 adults recruited via internet	High predictive accuracy of any depressive disorder (AUC = .84) High internal reliability (*N* = 502) Significantly different CES-D means by those with and without any depressive disorder	
	Fortson et al. (2006)	Equivalence with p-p Randomised order *N* = 411 university students	Similar internal consistency, means scores, and inter-item correlation between web and p-p format, across different time points – however, differences were not statistically tested.	

	Herrero & Meneses (2006)	Equivalence with p-p Randomised order *N* = 530 university students, frequent internet users	No significant differences in internal consistency No significant differences in factor structure based on non-significant chi square test	
	Yu & Yu (2007)	Equivalence with p-p Randomised order *N* = 2400 Taiwanese teachers	No significant differences in factor structures Small differences in latent mean	
Edinburgh Depression Scale (EDS) (self-rated scale)	Spek et al. 2008	Criterion validity (web BDI and web SCL-90) *N* = 407 participants with subthreshold depression	Good correlation with BDI (*r* = .72) and SCL-90 (.77-.72) Significant difference in mean EDS score between those with and without CIDI depression High internal reliability	Comparable psychometrics with p-p format for depression population
Montgomery-Åsberg Depression Rating Scale – Self Rated (MADRS-S)	Carlbring et al. (2007)	Equivalence with p-p Randomised order *N* = 494 applicants of online panic dis. intervention program	Generally equivalent mean scores and internal consistency High correlation between web and p-p (*r* = .91) Negligible difference in internal consistency	Strong equivalence with p-p format for panic disorder population
	Holländare et al. (2010)	Equivalence with p-p Randomised order *N* = 87 Swedish participants in primary/psychiatric	Similar internal consistency High correlation between web and p-p (*r* = .84)	
Major Depression Inventory (MDI)	Cuijpers et al. (2008)	Criterion validity (MINI) *N* = 243 adolescents	High internal consistency (.88) High correlation with web CES-D (*r* = .88) Poor sensitivity (.57), high specificity (.91) for MDD; moderate diagnostic agree. (kappa = .41) High predictive accuracy of MDD (AUC = .89)	Good validity and reliability, suitable for screening in adolescents
Hospital Anxiety and Depression Scale – Depression (HADS-D)	Andersson et al. (2003)	Equivalence with p-p Between subjects design *n* = 157 patients with tinnitus completed web HADS *n* = 86 completed p-p HADS	Higher percentage of clinical depression in internet HADS (17%) than p-p HADS (15%) Similar Cronbach alpha between web and p-p HADS	Adequate as screener for MDD when cutoff of 6 or greater is used
	Bunevicius et al. (2007)	Criterion validity (MINI) *N* = 503 primary care patients	When HADS cutoff equal or greater than 6: AUC = .75 Sensitivity = .80 specificity = .69	Some evidence of equivalence with p-p
Clinically Useful Depression Outcome Scale (CUDOS)	Zimmerman & Martinez (2012)	Equivalence with p-p *N* = 53 outpatients with depression	High corr. between web and p-p (*p* < .001) High and similar internal consistency (both Chronbach alpha = .93) Web and p-p versions had equal correlation with clinician rated scales (MADRS, Clinician Global Inventory-Scale, and GAF)	Strong equivalence with p-p format for outpatents with depression

Mobility Inventory (MI).	Austin et al. (2006)	Equivalence with p-p Randomised order *N* = 110 people with panic disorder	Similar internal consistency High corr. between web and p-p (*r* = .95-.96) No significant differences in mean score between web and p-p	Generally equivalent with p-p format for population with panic disorder
	Carlbring et al. (2007)	Equivalence with p-p Randomised order *N* = 494 applicants of online panic disorder intervention, recruited online	High correlation between web and p-p (*r* = .95-.96) For MI (*Alone* version), significantly higher mean scores on web format, small effect size For MI (*Accompanied* version), no significant differences in mean scores of web and p-p	
Schizophrenia Outcomes Module 2	Chinman et al. (2004)	Equivalence with interviewer administration *N* = 90 patients with bipolar and/or schizophrenia	Similar internal consistency High correlation between web and interviewer scores for bipolar sample (*r* =.97) than schizophrenia sample (*r* = .84)	High reliability and concurrent validity with interviewer ad.
General Health Questionnaire (GHQ-28)	Vallejo et al. (2007)	Equivalence with p-p *N* = 100 psychology students in Spain	Similar Cronbach alpha between web and p-p Correlation of total score between p-p and web = .69; less correlation for subscales. Equivalent factor structure (four factors)	Generally equivalent with p-p format in non-clinical sample
Kessler-10 (K-10)	Donker et al. (2009)	Criterion validity (CIDI – any depressive disorder) *N* = 157 adults recruited via internet	High predictive accuracy for any depressive disorder (AUC = .81) Good optimal sensitivity (.69-.81) and specificity (.67-.79) High internal reliability *(N* = 502) Significantly different K-10 means for those with and without CIDI diagnosis	Good criterion validity and reliability for adult population with depression
Symptom Checklist 90 (SCL-90-R)	Vallejo et al. (2007)	Equivalence with p-p *N* = 100 psychology students in Spain	Similar Cronbach alpha between web and p-p Correlation between web and p-p across subscales ranged from .63 to .86 Mean scores significantly higher for p-p (small to medium effect size across subscales)	Fairly equivalent with p-p format in non-clinical sample
CIDI – Short Form (CIDI-SF)	Carlbring et al. (2002)	Criterion validity (SCID-IV and Panic Disorder CIDI module) *N* = 53 participants, Swedish, recruited from newspaper	Low diagnostic agreement (kappa < .40) Generally poor sensitivity (.25-1.00) Low to high specificity (.55-1.00) Poor discriminative validity according to SCID-IV and panic disorder CIDI module	Generally poor criterion validity

Note. N = total sample, *n* = sub-sample; "p-p" = original paper-pencil form of measure; AUC represents area under the ROC curve; * studies were identified through extensive, though non-exhaustive database searches and from references in previous research articles.

As an example of a study reporting general equivalence, Austin et al. (2006) compared the paper-pencil and web-administered formats of three commonly used panic-related

questionnaires: the Body Sensations Questionnaire (BSQ), Agoraphobic Cognitions Questionnaire (ACQ), and the Mobility Inventory (MI). Web-administered versions of these questionnaires were developed on the basis of being brief, easy to understand, and simple to respond to (Austin et al., 2006). Participants (54 resided in Sweden and 56 in Australia) were previously diagnosed with panic disorder and were registered to undertake an online treatment program for panic disorder. A randomised design was employed whereby participants completed both web and paper formats in a randomly assigned and counterbalanced order.

The results showed that the paper and web modes had similar psychometric properties for the three questionnaires. The internal consistencies were largely the same across countries and administration formats. The strength of relationship between the paper and web administrations, as measured by the intra-class correlation statistic in this study, ranged from strong (.73) in the case of BSQ, to very strong (.96) for MI. Repeated-measures ANOVA found no interactions to suggest any order effects across the three questionnaires. However, the paper-administration of the BSQ resulted in a higher mean score compared with the web administration (Austin et al., 2006), which contrasted with suggestions of the opposite occurring (e.g. Buchanan, 2003). While Austin et al.'s (2006) study concluded there is strong equivalence between the online and paper administration of the questionnaires examined, the authors cautioned against alternating administration formats at different time points within a treatment study, given the small but significant differences in scores between the paper and web administration of the BSQ.

Online CES-D.

Whereas many studies have examined psychometric properties of online clinical measures through demonstrating their equivalence, other studies have compared them against "gold standard" criterion measures. Further to evidence demonstrating equivalence between a web-based and paper administration of the CES-D (Herrero & Meneses, 2006), Cuijpers et al. (2008) evaluated the criterion validity of a web-administered CES-D against the MINI interview, as part of a wider psychometric study of online depression measures for adolescent use. Participants, aged 14 to 16 years, completed an online questionnaire including the CES-D and MDI. Approximately one week after, a subset of participants (*n* = 243) undertook a telephone based MINI interview. Results found a MINI diagnosis of any mood disorder associated with significantly higher CES-D scores. Receiver Operating Characteristic (ROC) curves were calculated using MINI diagnosis as the validity criterion and indicated excellent diagnostic accuracy for the online CES-D (AUC = .90; 95% CI: .84 -.95) and MDI (AUC = .89; 95% CI: .82-

.95). ROC curve analysis also suggested optimal cut-off thresholds: a CES-D score of 22 indicated a sensitivity of .90 and specificity of .74. The authors noted this threshold was much higher than recommended for the paper-pencil CES-D and suggested potential sampling factors to explain differences (Cuijpers et al., 2008). In addition, the study found the CES-D displayed satisfactory internal reliability (.93) while CES-D scores significantly correlated with MDI scores (r = .88, p < .001). The study concluded that a web-based CES-D can be a valid and feasible means of screening depression in adolescents.

Online Edinburgh Depression Scale.

The Edinburgh Depression Scale (EDS) has also been validated for online usage against a structured interview (Spek et al., 2008). The EDS is a self-report scale comprising ten items with evidence of good psychometric properties (Spek et al., 2008). In Spek et al.'s (2008) study, 407 participants with a mean age of 55 years and subthreshold depression (indicated by cut-off score of 12 on internet-based EDS) were recruited from a trial of internet-based CBT. Participants also completed other web-administered depression scales and an in-person structured interview using the CIDI MDD module as validity criteria. Participants with a positive CIDI diagnosis of MDD were found to have a mean internet-EDS score of 20.18 (SD = 3.55), whilst those with an absence of diagnosis had a mean score of 16.48 (SD = 3.45). The internet-EDS was also found to significantly (p < .001) correlate with the internet-BDI (r = .75) and internet SCL-90: depression (r = .77). While these results indicate good validity, further validation was recommended with younger and less symptomatic population groups to generalise findings (Spek et al., 2008).

Internet CIDI-SF.

Some studies of criterion validity raise significant doubts about the evaluated program. Carlbring et al. (2002) evaluated the web-based Composite International Diagnostic Interview – Short Form (CIDI-SF) by comparing its results with an in-person administration of the SCID-IV. The authors proposed that an online CIDI-SF could reduce the time and financial burden of the assessment process in research trials. Participants (n = 53), recruited from an internet-based panic disorder treatment study, first completed the online CIDI-SF and the panic disorder module from the full CIDI. Approximately two days later, participants completed an in-person SCID-IV interview and three questionnaires: Mobile Inventory for Agoraphobia (MI), Anxiety Sensitivity Index (ASI), and Cardiac Anxiety Questionnaire (CAQ). SCID interviews were conducted by trained research assistants blind to the computerised interview, with interviews recorded to

assess for inter-rater reliability. Five random tapes were reassessed by an experienced clinician and the resulting diagnostic agreement was considered excellent (kappa = .81).

The results showed generally low agreement between the CIDI-SF and the SCID for the seven DSM-IV disorders considered (MDD and the main anxiety disorders). Cohen's kappa varied from .00 (MDD) to .39 (agoraphobia), with no disorders reaching the recommended cut-off for fair agreement (i.e., kappa = .40; Carlbring et al., 2002). Sensitivity varied between .25 (MDD) to 1.00 (OCD), and overall was considered low by the authors. Similarly, specificity ranged from .53 (GAD) to 1.00 (agoraphobia) and was also considered low. An analysis of variance was performed to see whether scores on the three questionnaires (i.e. MI, ASI, and CAQ) differed for those with and without a panic disorder according to the CIDI-SF, panic disorder module (CIDI), and SCID-IV. Significantly higher questionnaire scores were observed for those with a panic disorder diagnosis according to the full panic disorder module and SCID-IV, but not the CIDI-SF.

To explain the poor results of the online CIDI-SF, Carlbring et al. (2002) suggested the original CIDI-SF may have questionable validity given its limited psychometric evidence. It was also suggested that results showing CIDI-SF's higher percentage of diagnosis compared with the SCID-IV may be due to participants responding differently online (e.g., more honest disclosure) and because the CIDI-SF does not address diagnostic overlap (e.g., collapsing two overlapping diagnoses into one disorder), while the SCID-IV does. The authors raised the possibility that SCID-IV interviews may have lacked precision, though its validity was supported by its association with anxiety questionnaire scores. Nevertheless, it was suggested that a validity criterion such as the SCID may not be optimal, and that the LEAD standard (Spitzer, 1983) be considered in criterion validity studies of online assessment programs. However, Carlbring et al.'s study employed a small and select sample, which likely compromised the reliability of the results and its generalisability. Overall, given its poor validity results, the web-administered CIDI-SF was summarised as inadequate for screening purposes in clinical trials (Carlbring et al., 2002).

2.4 Examples of recently evaluated web-based clinical diagnostic programs.

The majority of evaluated web-based clinical assessment programs have been based on existing self-report questionnaires or clinical interview schedules. However, a few programs have been developed specifically for web use and undergone psychometric evaluation. Given the close relationship between these programs and the online program of focus in this research (i.e. the e-PASS), the studies are described in detail as follows.

2.4.1 The WB-DAT.

The Web-Based Depression and Anxiety Test (WB-DAT; www.wb-dat.net/WB-DAT/) is a freely available online self-report questionnaire for current diagnoses of MDD and the anxiety disorders. The item content and branching rules draw from DSM-IV and ICD-10. According to its developers, the WB-DAT was designed to have high sensitivity and reasonable specificity in detecting and ruling out its targeted disorders (Farvolden, McBride, Bagby, & Ravitz, 2002). The WB-DAT first presents 11 screening questions corresponding to targeted disorders and, depending on responses, branches into further questions covering relevant diagnostic criteria. A diagnostic summary is generated from the responses, which respondents are encouraged by the program to share with a health professional (Farvolden et al., 2002). Consistent with its intended function as a screening tool, the WB-DAT reports as many diagnoses as relevant to the user's responses (Farvolden et al., 2002). For example, if several disorders are detected, they are all reported as equally important, and it is assumed a health professional upon reviewing the results would confirm the correct diagnosis (Farvolden et al., 2002).

Farvolden et al. (2002) assessed the criterion validity of the WB-DAT by measuring its agreement with the SCID-I/P structured interview. Participants consisted of 193 adults recruited from current depression and gambling related research projects at a mental health centre in Canada. Participation involved completing the WB-DAT then, following an unstated amount of time, an interview with psychology graduate students using the SCID-I/P. Within the sample, the average number of diagnoses was 0.99 (*SD* = 1.45) according to the WB-DAT, and 0.79 (*SD* = 1.17) according to the SCID-I/P. Approximately 40% of participants were diagnosed with one or more disorders by the WB-DAT or the SCID-I/P. Due to low base rates in the sample, disorder specific analyses were limited to MDD, panic disorder with/without agoraphobia, OCD, specific phobia, GAD, and PTSD.

The WB-DAT's diagnostic agreement with the SCID-I/P, as measured by Cohen's kappa, varied from .57 (panic disorder with/without agoraphobia), considered acceptable, to .70 (PTSD) and considered good by the authors (Farvolden et al., 2002). Sensitivity ranged from .63 (GAD) to .95 (PTSD), whilst specificity varied from .89 (MDD) to .97 (OCD). In general, sensitivity and specificity values were considered good to very good. Positive predictive values ranged from .52 (panic disorder with/without agoraphobia) to .75 (MDD) whilst negative predictive values were consistently high (.93 – .98). The efficiency, which measures overall accuracy and is calculated by dividing the number of correctly classified cases by the sample size, was found to be generally high (.89 – .96). Overall, results suggested the WB-DAT has strong agreement with the SCID-I/P

and is a highly accurate diagnostic measure for MDD and most of the anxiety disorders. However, further validation with a wider sample group was recommended as the WD-DAT aims to be used in primary care settings (Farvolden et al., 2002).

2.4.2 The ISP-D.

The Internet-based Self-assessment Program for Depression (ISP-D) is an online screening program for depression in Chinese language and publicly available on a Taiwanese mental health website (PsychPark; Lin et al., 2007). The ISP-D was developed by psychiatrists with the aim of being a concise and interactive self-administered tool for assessing three depression types: MDD, minor depressive disorder (MinD), and subsyndromal depressive symptoms (SSD). Symptoms are assessed with items adapted from the MINI (Sheehan et al., 1998) and a Taiwanese depression measure (Lin et al., 2007). There are 24 items in total where a minimum of nine items are presented due to branching rules. In a rare study examining both validity and reliability (test-retest), Lin et al. (2007) evaluated the ISP-D's performance with adult participants recruited voluntarily from the PsychPark website. Of the total sample (n = 579), 184 participants repeated the ISP-D up to four weeks later. Test-retest reliability, as measured by weighted kappa, was considered excellent when the ISP-D was repeated within two weeks, but dropped considerably (though still deemed "fair") when retesting occurred two weeks or later. For instance, kappa value for MDD dropped from .83 (within two weeks) to .45 (two to four weeks).

The criterion validity of the ISP-D was investigated with a sub-sample of 55 participants, who were significantly older and more likely employed than the remainder of the total sample. These participants underwent a face-to-face semi-structured clinical interview with a psychiatrist (involving the MINI) a mean of 5.7 days after completing the ISP-D. Mean completion time was 10.29 minutes for the ISP-D, compared with 22.4 minutes for the interview. Only two participants were diagnosed with MinD or SSD, hence these participants were considered non-MDD cases, and only MDD was subject to validity analysis. Using the clinical interview outcome as the criterion, the ISP-D's validity statistics were: .82 sensitivity; .73 specificity; .67 positive predictive value; and .86 negative predictive value. Socio-demographic characteristics were not significantly associated with diagnostic agreement (Lin et al., 2007).

Lin et al.'s results regarding the validity of the ISP-D's MDD diagnosis are in line with other studies (e.g. Farvolden et al., 2002) showing reasonable diagnostic accuracy of a web-based questionnaire. However, unlike Farvolden et al.'s study which recruited participants from clinical research projects, Lin et al.'s study recruited from an internet-using population, which is a

more ecologically valid representation of those who use the program. On the other hand, the self-selected sample largely consisted of young, single, and well educated women, which limits the generalisability of the findings to the wider population. The authors also suggested the self-selection process may have attracted more depressed people than expected from population prevalence rates. Hence, Lin et al. suggested further validation of the ISP-D with different population groups and larger sample sizes.

2.4.3 The WSQ.

The Web Screening Questionnaire for common mental disorders (WSQ) is an online program based on a screening questionnaire developed by Marks and colleagues (Donker et al., 2009). Consisting of 15 items, the WSQ was designed to be brief and simple to use in order to encourage completion (Donker et al., 2009). It targets a range of disorders (MDD, major anxiety disorders, alcohol abuse/dependence) and is suggested as a screening tool prior to consultation with a health professional or to direct respondents to appropriate online self-help resources (Donker et al., 2009). The WSQ underwent a process of testing and subsequent revision in order to enhance its validity. Participants consisted of 502 Dutch adults who were recruited online (Donker et al., 2009).

To validate the WSQ, participants underwent the WSQ followed by a phone interview using the CIDI interview schedule as criteria. Participants also completed 10 online questionnaires including well-regarded scales (e.g. CES-D, GAD-7, etc) corresponding to disorders targeted by the WSQ. The results showed that those with a positive WSQ screening result for a particular disorder had significantly higher scores on the corresponding questionnaire compared with those who had a negative screen. For instance, participants screened by the WSQ as having GAD received a mean score on the GAD-7 questionnaire of 13.6 (*SD* = 3.9) whilst those who had a negative screening result had a mean GAD-7 score of 5.5 (*SD* = 3.1). Classification statistics using the phone interview results as criteria found four WSQ subscales (GAD, OCD, specific phobia, and panic disorder) with sensitivity and/or specificity values below the pre-specified threshold levels (i.e. sensitivity of .70, and specificity of .40).

In order to improve validity, the WSQ items of the GAD, OCD, and panic disorder subscales were substituted with items from corresponding questionnaires identified by logistic regression as having better validity statistics. The validity for specific phobia could not be improved and therefore the WSQ item for this remained unchanged. Resulting sensitivity values of the WSQ subscales range from .60 (specific phobia) to 1.00 (agoraphobia), with most

subscales valued between .80 to .90. Specificity values were between .44 (panic disorder) to .77 (panic disorder with agoraphobia). The positive predictive values varied between .11 (PTSD) and .40 (social phobia), whilst the negative predictive values ranged from .87 to 1.00. Based on AUC statistics, the WSQ subscales of GAD, OCD, alcohol dependence, and panic disorder were found to be similar to the longer corresponding questionnaires in terms of overall diagnostic accuracy.

In summary, the WSQ has the potential to achieve high sensitivity (if substituting certain items), which is comparable to other web questionnaires. However, specificity was generally low and it was suggested this may have been partly due to the use of 6-month prevalence rates in the phone interview, which differed from the point (i.e., current) prevalence focus of the WSQ (Donker et al., 2009). The authors found that while the WSQ appears good at screening out negative results, it produces a high level of false positives, misidentifying participants as falsely having a disorder. Therefore it was recommended that individuals screening positive on the WSQ then undergo a clinical assessment with a more specific diagnostic instrument. While the study displayed promising results for the WSQ, there were several limitations as noted by Donker et al.. Specifically, they did not examine the inter-reliability of the interviews to measure interviewer performance, or effects of completion order of the WSQ and interview. Donker et al. also highlighted that results formed from their self-recruited sample may not be generalisable to the internet-using population. Donker et al. suggested that future research look at refining questions and diagnostic criteria to enhance clarity for respondents, and in doing so, reduce false positive rates, particularly for the WSQ's GAD, panic disorder, specific phobia, and PTSD subscales.

2.6 User background, usability, and experience of web assessment programs.

As reflected in the notion of clinical utility, there are additional factors of importance when evaluating an assessment instrument and its suitability for a situation. These could include who will likely use it, how easy it is to access, use and comprehend, and how people respond to using it. To help discuss these issues in the context of web-based clinical assessment, two important paradigms in the human-computer interaction (HCI) literature will be introduced, namely *usability* and *user experience* (Hassenzahl, 2008; Hassenzahl & Tractinsky, 2006; Tullis & Albert, 2008). According to the International Organisation for Standardisation (ISO), usability reflects the extent of which a product is used by its target users and perceived to be effective, efficient, and satisfying based on users' own goals (Goldberg et al., 2011).

While usability focuses on performance and satisfaction related to instrumental or goal-oriented needs, the user experience paradigm takes a more holistic view of user interaction, and

considers the fulfillment of more general human needs such as autonomy, personal growth, stimulation, and pleasure of use (Bargas-Avila & Hornbaek, 2011). These broad (and overlapping) paradigms encapsulate many aspects of interest and are typically referred to during the design, development, evaluation, and revision of computer and online systems. The following section will review the limited empirical evidence relating to the usability and user experience of online assessment. There is relatively more literature for computerised assessment which will also be considered in relation to web assessment.

2.6.1 Why access online assessment programs?

An important factor in evaluating the experience of a web-based program is considering the needs of the user and whether they are met. More broadly, this is reflected by the reasons people access a program. To date, only one study has specifically asked consumers why they accessed a web-based clinical questionnaire (Van Ameringen, Mancini, Simpson, & Patternson, 2010). In a pilot study, Van Ameringen et al. (2010) approached users of the MACSCREEN, an online screening program for high prevalence disorders and psychotic symptoms. The study was reportedly the first to focus on members of the general population seeking online mental health information. The mean age of participants (n = 302) was 35.2 years (SD = 13.9 years) while most were single (46%), had a university or college degree (62.4%), lived in an urban area (86.9%), and received previous treatment for emotional issues (55.6%).

The study found almost two thirds of participants accessed the MACSCREEN due to concerns of having anxiety problems while the remaining participants were “just curious” (Van Ameringen et al., 2010). Roughly one in five participants completed the program to confirm a diagnosis from a health professional, suggesting the program was perceived to offer a credible diagnosis for some people. When asked about sought after information, at least half were looking for current symptoms (54.6%) or a specific medical condition (50.7%) they thought they had. Almost a third of participants were querying for a friend or family member whilst approximately one in five participants were looking for treatment information (Van Ameringen et al., 2010). Interestingly, almost all MACSREEN users had access to a doctor for consultation and did not emphasise the screening program as more convenient than seeing a professional, yet still went about accessing the program. As the authors highlighted, further research is needed to investigate how consumers specifically use internet-derived diagnoses and health information.

2.6.2 Acceptability of and preference for computerised assessment.

Acceptability is a broad measure of usability and, in the context of assessment, represents the extent of which an instrument will be adopted. Acceptability can have motivation effects on performance (Weber, Fritze, Schneider, Kuhner, & Maurer, 2002) and influence an instrument's completion rate and accuracy (Robins, 1994). Clinicians were initially doubtful about the acceptance of computerised assessment, as there were concerns it would be misunderstood or used with difficulty by impaired clients (Newman et al., 1997), and ultimately interfere with the therapeutic relationship (Endicott & Spitzer, 1975, as cited in Hile & Adkins, 1997). However, various findings have suggested that computerised assessments can be well accepted across different settings and amongst consumers (e.g., Carr et al., 1981; Greist, Klein, Jefferson, & Getto, 1981) and clinicians (Angle et al., 1978).

For example, several studies have shown psychiatric patients are willing and able to be interviewed by a computer at admission (Angle et al., 1978; Carr et al., 1981; Greist et al., 1981). Petrie and Abell (1994) examined the experiences of 150 newly hospitalised patients with parasuicidal presentation and found almost the entire sample (97%) agreed to complete a computerised psychiatric interview. Furthermore, most participants preferred the computerised interview (52.3%) over an interview with a doctor (17.4%), with computer preference higher in patients with lower self-esteem and higher levels of suicidal ideation and hopelessness.

However, lower acceptance rates of computerised assessment have also been observed (Blouin et al., 1988; Erdman et al., 1992). For instance, Peters et al. (1998) found more patients at an anxiety clinic preferred an interviewer administered CIDI (45%) over the computerised version (25%) while remaining participants had equal preference. In Erdman et al.'s (1992) study comparing a computer and interviewer administered DIS, participants were generally positive towards the computer interview and found computer interaction less embarrassing. However, participants reported greater difficulty describing their feelings to a computer and, overall, had similar preferences between administration formats (Erdman et al., 1992).

Evidence also suggests acceptance for computerised assessment differs between population groups. Kobak, Reynolds, and Griest (1994) found psychiatric patients tended to prefer a clinician over a computer administered DIS, whereas the control group had no strong preferences. Another study found participants with lower reading ability had greater preference for computer as opposed to human administered interview (Erdman et al., 1992). The authors suggested that people with lower reading ability may prefer computerised assessment because it permits more time to process and respond to verbal questions (Erdman et al., 1992). In a study

involving older aged adults, Kurt et al. (2004) evaluated attitudes towards computerised depression measures amongst a large sample of adults over 65 years of age under primary care. Despite 72% of the sample indicating no previous computer use, there was a generally favourable response, with the vast majority indicating comfort and liking computer use. However, roughly half the sample found it frustrating at times, and a portion of the sample reported some computer-related anxiety prior to (28%) and/or during the program (19%).

Clinicians have also shown mixed reactions to computerised assessment (Angle et al., 1978). In an earlier study by Angle et al. (1978), the majority of clinicians viewed a computerisd psychiatric interview as more comprehensive, detailed and accurate in problem identification than clinician administered interviews. However, the program's utility in assessment and treatment formulation was rated more poorly by clinicians from a hospital psychiatric setting compared with a community mental health clinic. More recently, Chinman et al. (2007) found treating psychiatrists expressed a liking towards a web-based clinical measure used in an outpatient clinic setting because it validated patients' experiences and helped generate conversation with patients. However, interviewed psychiatrists also felt it lacked overall clinical utility and this was reflected in the low proportion of clinicians referring to assessment reports during consultation (Chinman et al., 2007).

2.6.3 Ease of use.

Ease of use is an important indicator of usability. Evidence supporting the ease of use of computerised programs for clinical assessment extends back to early applications. A study in 1981, for example, examined use of a computerised program for collecting personal history in an inpatient psychiatric setting (Carr et al., 1981). Results found 88% of the sample rated the computer program as easy to perform as undertaking an interview. More recently, developers have become more mindful of designing web-based programs that are easy to use by people with mental illness (Rotondi et al., 2007) and promote behaviour change (Danaher, McKay, & Seeley, 2005). However, programs are constantly evolving due to technological advances and the ambitions of developers. Moreover, the growth in internet use implies that programs are being accessed by people with varying abilities and preferences. Therefore, a challenge in web usability is maintaining ease of use while catering to a wide range of people.

While perhaps taken for granted by regular web users, online programs can involve complex cognitive activities. Website use in general is associated with spatial and verbal ability, as well as cognitive functions such as executive functioning, working memory, and sustained

attention (Rotondi et al., 2007). Web programs could therefore be harder to use for individuals with cognitive deficits, such as those associated with severe mental illness (Rotondi et al., 2007). Chinman and colleagues (2004) explored this by evaluating the usability of the web-based Patient Assessment System (PAS) with 90 outpatient adults diagnosed with bipolar disorder or schizophrenia. The PAS was specifically designed to minimise cognitive load (e.g., items presented in audio and text with clear wording, simple visual layout) and comprised of items from standardised symptom and functional measures (Chinman et al., 2004, 2007).

When compared to an interviewer administered PAS, the online version was completed in similar time, yet was overwhelmingly preferred and rated as "easier to take" than the interviewer administration (Chinman et al., 2004). Almost all participants could describe the program functions and less items were skipped in the online PAS compared to the interviewer administration (Chinman et al., 2004), suggesting that participants felt more able to respond to the online administration. In another study, an expanded PAS including items about various symptoms, substance use, medication, and side-effects, was tested for feasibility in an outpatient mental health clinic (Chinman et al., 2007). The sample comprised 266 patients with high prevalence and/or severe disorders, and 14 psychiatrists. Focus groups, interviewing, and surveying found the PAS to be easy to learn, comfortable, and even enjoyable to use by some. Psychiatrists also reported the program as easy to use, with no recollections of bad experiences. These results demonstrate that when program design addresses the cognitive needs of the target population, online assessment can be highly usable by people with severe mental disorders.

2.6.4 Comprehension.

Self-administered web-based assessment programs that are text-based place an emphasis on respondents' ability to accurately comprehend assessment instructions, items, response options, and feedback. As with all questionnaires, there is the risk that ambiguous content will be misinterpreted (Kessler & Ustun, 2004), though it is unclear how significant an issue this is in web-based assessment. An early systematic study of printed questionnaire surveys found over 70% of a sample interpreted questions differently to researchers' intentions (Belson, 1981, as cited in Kessler & Ustun, 2004). In studies where participants were debriefed about their survey experience, misinterpretation has also been found to be a common issue (Oksenberg, Cannell & Kanton, 1991; Kessler & Ustun, 2004).

Kessler et al. (1999) identified four survey item characteristics commonly associated with miscomprehension: items were too complex in terms of language or concepts; items were too

vaguely defined; items referred to odd experiences that could be easily misinterpreted; and context of items was not clear due to its position amongst other scheduled questions. With the lack of supervision to clarify queries, it is not surprising that some applications of computerised assessment have been rated by consumers as less easy to understand than a clinician administered interview (Blouin et al., 1988). The issue of miscomprehension could be particularly pronounced with web-based assessment accessed by people of different socio-cultural backgrounds who may differ in how they perceive and respond to content (Suler, 2001).

2.6.5 The cognitive effects of computerised questionnaires.

Earlier experimental comparisons of computer- and paper-based tasks suggested several potential effects of computer administration on cognitive processes and outcomes (see Dillon, 1992, for a seminal critical review). Firstly, it was reported that reading was slower on a computer screen than on paper with reported effects varying from minimal (e.g. Keenan, 1984) to up to 20 to 30% slower (Dillon, 1994). Reading accuracy was also found to be better on paper than on computer-based text (Creed et al., 1987), though others noted no significant difference (Askwall, 1985). Studies of comprehension showed participants detected twice as much information (Askwall, 1985) and solved problems faster (Weldon et al., 1985) on paper-based materials as opposed to computer related media.

Some suggested these unexpected findings were likely due to differences in visual quality between presentation formats (Noyes & Garland, 2008). Along these lines, Ziefle (1998) concluded that paper previously performed better than computer presentation because of the poor display screen quality found in computers during the 1980s and early 1990s which caused visual (and cognitive) fatigue. However, more recent studies involving better display technology continued to show inconsistencies; some found the computer medium to be associated with relatively slower reading (Mayes et al., 2001) and poorer comprehension (Wastlund et al., 2005), while others reported no differences between computer and paper forms (Mayes et al., 2001, Noyes & Garland, 2003).

Drawing on previous findings and theories, Noyes and Garland (2008) identified several technological characteristics that could influence cognitive performance and therefore overall computer-based assessment performance. For instance, the visual properties of a computer monitor (e.g., flicker, high contrast, brightness) were suggested to affect how information is processed and retrieved from memory. In particular, the computer medium was purported to require perceptual and executive cognitive resources that resulted in a higher cognitive workload,

tiredness and feelings of stress than associated with the paper medium. Noyes et al. (2004) investigated this hypothesis and found that while comprehension scores did not significantly differ between mediums, the computer-based task involved significantly more work load on the effort dimension of a cognitive workload measure (i.e., NASA-Task Load Index) than the paper-based task. In summary, Noyes and Garland concluded that the computer medium can affect cognitive workload for respondents and create small but potentially important variation in performance, especially in more complex tasks requiring sustained concentration, problem solving and so on.

Bowling (2005) builds on this notion by proposing a model outlining a range of cognitive and behavioural processes that could occur during assessment performance. For example, a respondent completing a self-administered questionnaire needs to comprehend the material, recall required information from memory; evaluate the relevance of recalled information to the question; and communicate the response (Bowling, 2005). According to Bowling, the cognitive demand underlying these processes varies between different modes of assessment, and ultimately impacts on data quality. A face-to-face interview is regarded as the least cognitively demanding mode as it requires only basic verbal and listening skills, and does not involve tasks such as writing or reading (Bowling, 2005). In contrast, Bowling argues that paper-pencil self-administered questionnaires represent the most demanding mode as they require literacy, visual functioning, physical dexterity, and the ability to follow routine instructions (Bowling, 2005). The tasks, cognitive processes, and workload likely differ with web assessment (depending on the program), though it is unclear how much impact this has on perceived experience and assessment outcomes.

2.6.6 Previous computer experience and diagnostic accuracy.

There has been suggestion that computer experience could affect performance on a computerised clinical assessment program. A lack of computer literacy, for example, might interfere with navigating and responding on a computer. However, reported studies have shown previous computer experience has little impact at least on the accuracy of certain computerised assessment programs. Peters et al (1998) found that the majority of participants (patients at an anxiety clinic and general practice) in their study were comfortable with a computerised CIDI interview (CIDI-Auto), and that previous computer experience had no significant bearing on diagnostic discrepancy between the CIDI-Auto and an interviewer administered CIDI.

In a study examining the online and paper-pencil equivalence of two OCD scales, computer use and experience was measured with the 12-item CUE scale (Coles et al., 2007).

Participants, comprising of university students, were generally comfortable with computers according to mean CUE ratings. The correlation between CUE and OCD scale scores were not significant, suggesting computer experience had no influence on how the OCD scales were rated (Coles et al., 2007). The results of these studies suggest most people are comfortable with computer use, and that computer use and experience has no major bearing on diagnostic accuracy of computerised/online measures. However, it remains uncertain as to whether variations in previous computer experience impacts on the overall usability and experience of an online assessment program.

2.6.7 Attitudes.

A person's attitudes towards computers and the internet could influence how people interact with web-based clinical assessment (Schulenberg & Yutrzenka, 2004). Weber et al. (2002) examined the influence of computer attitudes on the completion of computerised assessment within an inpatient setting. Seventy eight participants underwent clinical assessment, conventional and computerised memory and attention cognition tasks, and a self-report computer attitudes questionnaire. A more negative attitude towards computers was found to significantly correlate with higher levels of psychopathology and lower scores on attentional tasks. For those with depression, computer attitudes accounted for 39% of the variance in attentional performance, indicating that lower attention levels more likely associated with poorer computer attitudes (Webber et al., 2002).

Peters et al. (1998) investigated whether computer attitudes and computer experience predicted the level of diagnostic agreement between a self-administered CIDI-Auto and an interviewer administered CIDI. Computer attitudes were assessed using items taken from previously developed scales and addressed general attitudes (e.g.,"I feel a computer is too impersonal to give information to"), perceived computing ability, and computer-related anxiety (Peters et al., 1998). Scores on the computer attitude items significantly predicted the number of diagnostic discrepancies between the standard CIDI and CIDI-Auto, whilst other measured factors (e.g., computer experience, social desirability) did not (Peters et al., 1998).

2.6.8 Anonymity.

Anonymity is often associated with the online experience. Anonymity is traditionally regarded as the inability of others to identify an individual or their sense of self (Christopherson, 2007). Two broad types of anonymity have been defined (Hayne & Rice, 1997): *technical*

anonymity refers to the absence of meaningful identifying information (e.g., name) from communication; and *social* anonymity, which refers to perceived sense of being unidentifiable due to the lack of cues which attribute an identity to an individual. The online experience can offer both technical and social anonymity, for example, by removing visual and auditory cues, as well as allowing online use without directly revealing one's personal details (Christopherson, 2007). In doing so, anonymity can have powerful effects on computer related interaction.

Several theories have been proposed to explain the effect of anonymity on online behaviour. The equalisation hypothesis (Dubrovsky et al., 1991, as cited in Christopherson, 2007) suggests the lack of visual cues during computer use can free individuals from having to act according to social roles typical of their group membership (Postmes & Spears, 2002) and elicit a sense of empowerment amongst those who might traditionally have less power in society (Christopherson, 2007). Zimbardo's (1969) deindividuation theory refers to how anonymity and environmental factors creates a deindividuated state which in turn reduces self-observation, concern for social evaluation, and ultimately self-regulation. The deindividualised state is argued to weaken internal controls (e.g. feelings of guilt, fear, and shame), causing disinhibited and potentially hostile behaviour (Zimbardo, 1969; Kiesler et al., 1984). Given mixed support for Zimbardo's original theory, Prentice-Dunn and Rogers (1982) reconceptualised the deindividualised state as a combination of experiencing reduced accountability of one's actions; and decreased private self-awareness (i.e., of one's behaviour).

Alternatively, the online influence on behaviour has also been framed in terms of enhancing self-focus through changes in private and public self-awareness (Joinson, 1999). Drawing on the research of Matheson and Zanna (1988), it is purported that by lowering attention to oneself as a social object (i.e., public self-awareness), the anonymity associated with online use allows an individual to divert their focus to self-internal states (e.g., thoughts and feelings) and standards (i.e., private self-awareness; Joinson, 1999; Pinsonneault & Heppel, 1998). Unlike previous deindividuation theories, this explanation suggests anonymity enhances private self-awareness and self-regulation. In doing so, assessment conducted online and with anonymity has the potential to elicit more self-honest and open disclosure of issues, therefore revealing the "real" self of respondents (Barak & Hen, 2008; Joinson, 1999).

Privacy.

According to Christopherson (2007), anonymity has a positive influence on wellbeing by enabling privacy. Privacy involves removing ones presence from a social context, and can simply

refer to the ability to maintain boundaries which prevents others from accessing one's self (Pedersen, 1997). In a comprehensive factor analytic study, Pedersen (1997) identified three main psychological functions of privacy in the context of anonymity: recovery, autonomy, and catharsis. Recovery was the main factor and was defined as a type of rejuvenation involving contemplation of one's situation and a sense of relaxation and refuge from others. Autonomy, according to Pedersen, described the opportunity privacy affords for engaging with new behaviours without fear of being judged. Catharsis referred to the unrestricted expression of thoughts and feelings to others (Pedersen, 1997), and could have both positive and negative consequences (Christopherson, 2006).

Other functions of privacy identified by Pedersen (1997) likely also relevant to the online experience are *solitude* and *isolation*. In the aforementioned study, Pedersen (1997) found the functions of solitude and isolation comprised a process of contemplation and self-discovery which involved considering who one is and wanted to be. Autonomy was also relevant to both functions and, similar to its relationship to anonymity, referred to the opportunity to undertake new behaviour. Rejuvenation was also found to relate to solitude and isolation. Pedersen suggested that rejuvenation reflected people's experiences of social hurt, leading to social withdrawal (i.e., solitude/isolation) and a need to recover from the hurt and plan for future social interaction. While these findings refer to privacy needs and behaviour in general, they could be relevant to the needs and experiences of people using web-based assessment.

Reduced social desirability.

By providing a sense of anonymity, the internet experience has been shown to reduce social desirability (Joinson, 1999). Social desirability describes the intention of responding to a social situation in accordance to social norms, or in a manner believed to be socially desirable, rather than how one would act independently (Maccoby & Maccoby, 1954). In an assessment scenario, social desirability distortion describes a response pattern that is intended to be socially desirable (Richman, Keisler, Weisband, & Drasgow, 1999). Typically, intentional impression management is the main process underlying social desirability distortion, which in testing scenarios includes deliberately "faking bad" to gain an outcome (e.g., compensation, sympathy), or "faking good" to hide sensitive information or elicit a positive impression (Richman et al., 1999).

Reduced social desirability has long been recognised as an outcome of computerised assessment (Greist & Klein, 1981; Greist et al., 1987; Richman et al., 1999). Richman and colleagues (1999) conducted a meta-analysis of 61 studies comparing the social desirability of

computer versions of paper-pencil questionnaires and face-to-face interviews, with the original format. In addition to looking at discrepancies between results, they also factored measures of social desirability such as the K (Defensiveness) and L (Lie) scale on the MMPI, the Social Desirability Scale and the Crowne Marlowe Need for Approval Scale (Richman et al., 1999). The computer format was associated with the least amount of social-desirability distortion, especially when respondents were alone and could backtrack through their responses, and when interviews concerned highly sensitive personal behaviour such as the use of illicit substances and risky sexual behaviour (Richman et al., 1999).

More recently, Grieve and Elliot (2013) examined the intentions to fake in online psychological tests in comparison to faking in paper-pencil tests. Participants responded to an online questionnaire measuring attitudes towards faking, perceived behavioural control in faking, and intentions to fake in future "psychological testing". Results indicated participants held more positive views of, and felt more control over faking online, and these factors associated with intentions to fake (Grieve & Elliot, 2013). However, intentions to fake in future testing were ultimately found to be similar towards the online and paper formats. While these mixed results do not discount the vulnerability of online testing, further research is needed to clarify whether the online format promotes actual faking in psychological testing.

In relation to social desirability, Grieve and de Groot (2011) addressed the interesting question of whether the web format would facilitate test-faking any differently to the paper-pencil format when participants were prompted to fake-bad. Participants, consisting of an online recruited sample of university students and the general population, were found to produce similar results on both formats when asked to fake depression on the DASS-21 measure. The authors concluded that the online and paper formats were generally equivalent when respondents engaged in faking bad, though it was noted that the study did not address the issue of whether the web format increased the tendency to fake-bad (Grieve & de Groot, 2011).

Online disinhibition.

Reduced social desirability is also linked to why internet use is associated with disinhibited behaviour (Barak & Hen, 2008). The online disinhibition effect describes how individuals act differently online, in a manner that is normally constrained in day-to-day settings (Suler, 2004). Online disinhibition is often manifested in negative behaviour such as the expression of aggression, manipulation, and impersonation. In the context of online clinical assessment, disinhibition could for example result in a more reckless approach to responding, or

other assessment interfering behaviours in part influenced by underlying psychopathology. However, online dishinhibition could also lead to positive behaviour (e.g., offering support, donation, enhanced self-awareness, and self-expression; Barak & Hen 2008). For example, reduced inhibition may lead a normally reserved individual to engage in online socialising. A product of disinhibition and reduced social desirability that is particularly relevant to clinical assessment is enhanced self-disclosure.

Self-disclosure.

Self-disclosure refers to the communication of personal information to others (Archer, 1980). Early research into online self-disclosure examined how it leads to the development of new and meaningful relationships with others (Rheingold, 1993). As Bargh, McKenna, and Fitzsimons (2002) explained, this is because the relative anonymity afforded by typical internet use enables an individual to express themselves without the social constraints placed on them in a more public context. Hence, with anonymity online, the personal costs and risks of disclosing socially sensitive content are greatly reduced (Weisband & Kiesler, 1996). Given its influence on self-disclosure, the online environment has been argued as an ideal setting for conducting assessment (Barak & Hen, 2008; Bargh, McKenna, & Fitzsimons, 2002).

Computer use in itself has been found to be sufficient in enhancing self-disclosure in clinical measures (Weisband & Kiesler, 1996). An early study by Lucas et al. (1977) investigated differences between a computer and psychiatrist administered interview of patients with alcohol problems. While there was general agreement in content, participants reported significantly greater amounts of alcohol use to the computer as opposed to psychiatrists (Lucas et al., 1977), a result that has been supported in subsequent studies (Bernadt et al., 1989). As one would expect then, web-based questionnaires have also been found to facilitate self-disclosure. Kays, Gathercoal, and Buhrow (2011) examined rates of missing data to health related questions presented in both web and paper format and completed by university students. The study found questions independently rated as relatively sensitive (i.e., substance use, mental health, sexual practice) were less responded to on the paper-based survey compared to the web-format, suggesting the web enhanced self-disclosure over the paper-pencil format.

However, the relationship between administration format, social desirability, and self-disclosure may not be straightforward. In a unique study, Joinson (1999) investigated the effect of anonymity on disinhibition and disclosure by having participants either anonymously or non-anonymously complete measures of self-esteem, social desirability, and social anxiety in a web

or paper based questionnaire. Participants who completed the online version under anonymity scored significantly higher on self-esteem and lower on social anxiety and social desirability, compared to those who completed the paper-pencil version and without anonymity. Interestingly, those who were anonymous performed similarly to those who were non-anonymous in the online condition, suggesting that the non-anonymous condition was nullified by the effect of web-use on reducing public awareness and disinhibition (Joinson, 1999).

Factors contributing to anonymity, disinhibition, and self-disclosure.

Several factors could influence the level of anonymity and subsequent social desirability, disinhibition, and self-disclosure experienced by computer and internet users (Barak & Hen, 2008). For instance, the use of plain text, explicit instructions stating that responses will be anonymous, and fewer cues or reminders of the social context of the assessment might enhance a respondent's perceived anonymity (Barak & Hen, 2008; Joinson, 2001; Sproull & Kiesler, 1991). Furthermore, the social presence of others (e.g., photo of researcher) has been purported to reduce the willingness to respond to sensitive questions (Tourangeau, Couper, & Steiger, 2003). Conversely, programs that notify respondents that their responses will be identified, verified, and stored could compromise anonymity (Rosenfeld, Booth-Kewley, Edwards, & Thomas, 1996). There is also suggestion that programs that constrain the response process (e.g., not allowing the option to skip items, change previous responses, or select "don't know") can make respondents feel more self-conscious and cautious, and therefore respond with greater inhibition and social desirability (Hanna et al., 2005; Richman et al., 1999).

2.6.9 The emotional effect of web-based clinical assessment.

Some computerised assessment programs have been linked with positive personal experiences. For example, participants in a usability evaluation of a web-based assessment program in an outpatient setting on average rated it as an enjoyable experience, with some cases reporting they were proud of their reports (Chinman et al., 2007). These results suggest that an online assessment program can elicit a positive user experience, especially when the program is specifically designed to meet the needs of its users (Chinman et al., 2007). Not surprisingly, given the reduced social presence, computerised assessment is also associated with less embarrassment than interviewer-administered assessment (Erdman et al., 1992; Kobak et al., 1994; Peters et al., 1998). Peters et al. (1998), for example, found that within a sample of anxiety clinic patients, a larger proportion (38%) reported feeling more embarrassed by an interviewer CIDI than those reporting the computerised CIDI as more embarrassing (3.8%).

On the other hand, there are concerns that web-based clinical assessment programs have potential to cause negative emotional reactions amongst some individuals, particular those who are psychological vulnerable (Barak & Buchanan, 2004; Buchanan, 2002; Buchanan & Smith, 1999). Such concerns extend on those previously expressed for traditional assessment methods. Scarvalone and colleagues (1996) for instance suggested that clinical interviews could be overwhelming and distressing because of the large quantity and sensitive nature of clinical items (Scarvalone et al., 1996). Evans (2002) also noted that clinical questionnaires could result in harm because they elicit uncertainty (especially when lacking results) and often lack adequate support. In particular, clinical questionnaires can raise unnecessary worry in those without the actual problem or worsen existing concerns in those with the actual problem (Evans, 2002). Harm may also arise when a questionnaire raises positive expectations (e.g., the absence of a problem) only to be contradicted by further assessment (Evans et al. 2002).

Evidence indicates traditional forms of assessment can have a negative emotional effect. For instance, several studies have found mental health surveys elicit negative emotions, though only within a very small proportion of people (Henderson & Jorm, 1990; Jorm et al., 1994; Jacomb et al., 1999). For example, Jorm and colleagues (1994) examined the emotional impact of a mental health survey on a large sample of elderly adults regarding depression, cognitive decline, and life circumstances. While the majority reported no adverse effects, a small percentage of the sample found it distressing (4%) or depressing (1%). The researchers suggested that distress may have resulted from poor performance on cognitive testing involved in the survey (Jorm et al., 1994).

With a more diverse adult sample, Jacomb et al. (1999) looked at the emotional response of a mental health survey conducted by an interviewer in respondents' homes. Consistent with previous studies, a small number of participants reported feeling distressed during (5%) and following (2.5%) the interview. The authors examined predictors of emotionality and found that those with higher negative emotionality were more likely to report distress, while those higher in positive emotionality more likely reported a positive reaction to the interview (Jacomb et al., 1999). These results resembled Henderson and Jorm's (1990) results, which found that reported distress appeared related to pre-existing neurotic symptoms and personality issues (i.e. neuroticism) indicating a vulnerability to distress.

A limitation of these studies was that they retrospectively measured people's experiences. Scarvalone et al. (1996) took an alternative approach of measuring distress before and after the SCID-III-R structured interview (Scarvalone et al., 1996). The study found 54% and

72% of participants (individuals tested/treated for HIV in research trials) experienced a reduction in depression and anxiety symptoms, respectively, according to the Visual Analogue Scale (VAS). This was interpreted as suggesting the SCID-III-R reduced rather than increased distress (Scarvalone et al., 1996). However, questions were raised about the duration of the effect (e.g., whether reduction was transient) and whether distress was initially elevated in response to being in a novel setting (i.e. interviewed by a stranger), and lowered as respondents settled into the assessment (Scarvalone et al., 1996).

2.6.10 The impact of online assessment.

An important factor of an online assessment program's utility is what consumers gain from it. By asking specific questions about symptoms and providing assessment results, online assessment could facilitate self-reflection, raise people's awareness of their mental health, and subsequently influence follow-up behaviour (Barak & Buchanan, 2004). The potential positive impact of an assessment is central to the paradigm of *therapeutic assessment*, which purports that the administration of tests and provision of personal feedback can alone result in psychological improvement (e.g. reducing symptoms; increasing hope and goal focus; Finn, 1992; Finn & Tonsager, 1997). More specifically, the empirical assessment literature indicates that undergoing clinical interview schedules and questionnaires can have an effect in raising awareness of mental health issues. For example, the MINI interview schedule has been associated with promoting insight by reminding people about their symptoms (Pinninti et al., 2003). Also, a publically available computerised screening has been shown to enhance the awareness of mental health problems in roughly 50% of those screened (Ogles, France, Lunnen, Bell, & Goldfarb, 1998).

However, concerns have also been raised about the negative influence of web-based assessment feedback on consumers, particularly when results may be inaccurate (Buchanan, 2003; Barak & Buchanan, 2004; Naglieri et al., 2004). While web-based tests are often accompanied with disclaimers about their limited accuracy and the need to seek formal consultation, it is questionable how respondents react to their results. In relation to this, Barak and Buchanan (2004) draw attention to the "Barnum effect" in suggesting that people may come to accept web-based test results of high-base rate issues even when it is inaccurate for the individual.

There are significant effects of diagnostic labelling that could be relevant to the experience of web-based programs presenting diagnostic results. For instance, Hayne (2003)

conducted a qualitative study and found that a diagnosis of mental illness can present as abrupt and inescapable news for an individual, which can both legitimise and de-legitimise aspects of one self. Hayne also found that a diagnosis could make "visible the invisible" (p. 26) in terms of elucidating one's past experiences, current illness, and future direction, which ultimately were life-benefitting. Hayne noted how knowledge of diagnosis was helpful when it resulted in clients feeling better informed in how to improve their functioning.

Pitt, Kilbride, Nothard, and Morrison (2009) similarly found both positive and negative elements of the impact of psychiatric diagnosis. While a diagnosis can help "name the problem" and eventually facilitate a means of access to support and treatment, it can also result in social exclusion, a sense of disempowerment, and label the individual in negative terms (Pitt et al., 2009). While relevant to those who have received a formal diagnosis, these themes may not necessarily be experienced by those who receive diagnostic feedback from a web-based assessment program. As Barak and Buchanan (2004) caution, the personal growth that can come from online assessment depends on the validity of the instrument and the consumers' "effective assimilation of the meaning and implications of the results" (p. 221).

2.6.11 The follow-up actions of web-based assessment.

Many people go online for mental health related resources and are influenced to some extent in terms of follow-up actions and outcomes. In Lam-Po-Tang and McKay's (2010) qualitative study, psychiatric outpatients described how collecting online information improved their understanding of their condition as well as influenced their treatment choice and interactions with doctors. However, there are a few studies that have examined the effectiveness of web-based assessment programs in promoting positive outcomes for the consumer. One area where online assessment has been shown to be beneficial is in problematic alcohol use.

There is consistent evidence showing personalised feedback (e.g. normalised alcohol consumption levels, associated health risks) can lead to modest effect sizes in reducing alcohol consumption (Riper et al., 2009), and this finding has been extended to web-based programs functioning in a similar manner (Riper et al., 2011). For instance, Cunningham et al. (2006) evaluated an online assessment feedback intervention program for program drinkers. Compared with moderate drinkers, current problem drinkers in the sample found the feedback report more useful, surprising, and accurate (Cunningham et al., 2007). In addition, a significant reduction in various drinking measures was observed three months after participants received the online report, though a control group was lacking to confirm this effect (Cunningham et al., 2007).

A potential mediating factor in whether online assessment will lead to positive change is if it influences help-seeking behaviour and facilitates treatment management. There is some evidence to suggest that a computerised assessment result directly encourages patients to communicate with their healthcare professionals (Chinman et al., 2007). When computerised assessment results are made available to treating psychiatrists, the results have been described by psychiatrists as helpful in generating conversation, prompting questions about symptoms that may have been forgotten, and narrowing the areas of questioning to fit short appointment times (Chinman et al., 2007).

Only one study has so far evaluated the follow-up effects of a web-based screening questionnaire on a potentially severe population (Parker et al., 2013). Given the innovation of this study and its relevance to the present research, the background of the study and its findings will be described in detail. The program in question is a web-administration of the Mood Swings Questionnaire (MSQ), a 27 item checklist (rated on a 3-point scale) developed by Parker and colleagues (2009) for the screening of bipolar disorder. The MSQ is publically available on a university clinic website and is advertised as a “Bipolar Self-Test”.

Parker et al. (2013) recently investigated the follow-up actions taken by Australian adults (n = 665) who had completed the MSQ three months prior. In a novel approach, Parker et al. focused on participants who received a positive MSQ result but did not have a pre-existing diagnosis, and grouped them according to whether they: 1) did not take action; 2) took action (and either did or did not seek consultation), and; 3) took action and confirmed their diagnosis with a health professional. Only 16% reported not taking any subsequent action, with the most common reasons being busy, not wanting to take medication, and being unable to organise health professional consultation. Some were also open to the idea of having bipolar disorder but either required a second opinion before taking action or did not believe treatment was required.

Otherwise, remaining participants (84%) who received a positive screen but did not have a previous diagnosis indicated taking some form of action. This most commonly referred to seeking relevant information (e.g., books, internet sites), commencing or maintaining mood disorder medication, and starting or increasing exercise. A small subset of this group did not seek diagnostic clarification for various reasons such as not being able to afford consultation or because they viewed the MSQ as sufficiently accurate as to not require a formal diagnosis. The remaining subset (67%) sought diagnostic clarification, most commonly with a GP or psychiatrist.

Perhaps indicating the diagnostic accuracy of the MSQ, only a third of those who sought consultation had their positive MSQ result confirmed as correct. Health professionals either

"disagreed" or were "unsure" of the remaining two thirds of participants, though approximately 40% were subsequently referred to a specialist for a second opinion. Those who received a formal diagnosis from a health professional were also asked about their reactions. While the vast majority indicated a positive response (e.g. felt relieved, more control), negative reactions such as feeling upset (47%), angry (26%) and "in shock" (28%) were also evident, which pointed towards the negative effects of diagnosis previously mentioned.

A caveat when interpreting Parker et al.'s (2013) findings is that the MSQ, although implied in the study, may not have necessarily played a significant causal role in influencing follow-up actions. As participants were seeking bipolar testing, it is likely they had existing treatment motivation. Also, other resources (e.g., other websites, concurrent health professional) accessed during the three month follow-up period may have also prompted action. However, Parker et al.'s results suggest promise of the role of web-based screening in producing positive outcomes, especially in terms of encouraging personal research, lifestyle changes, and health professional consultation.

2.6.12 Trust and credibility.

Trust may be another important factor which influences engagement with and disclosure to an online clinical assessment program (Joinson & Paine-Schofield, 2007). Trust can be defined as the willingness to be vulnerable to another party based on positive expectations of their intentions (Rousseau, Sitkin, Burt, & Camerer, 1998). Concerns about one's privacy (e.g., personal data will be stolen or misused) can underlie a lack of trust towards a program and subsequently diminish self-disclosure and interaction (Joinson & Paine Schofield, 2008). Another aspect of trust relates to how people perceive and respond to web-based assessment results. Survey results of peoples' trust in the internet as a credible source of mental health information have shown mixed results.

In a population survey, Powell and Clarke (2006) found only 12.1% of their sample viewed the internet as the most accurate source of mental health related information, and was ranged eighth behind other sources such as a mental health professional (59.7%), "leaflets" from health related organisations (29.8%), and someone else with the same problem (15.3%). On the other hand, Leach, Christensen, Griffiths, Jorm, and Mackinnon (2007) found that at least half of their sample (comprising Australian adults) viewed online mental health websites has helpful, though there was generally greater trust in face-to-face contact (e.g., from a health provider), especially by older adults. In Lam-Po-Tang and McKay's (2010) survey of psychiatric outpatients,

most rated online mental health information as “very reliable” (20%) or “fairly reliable” (66%), while few rated “neither reliable or unreliable” (12%) or “unreliable” (1%). The authors suggested that participants in their study had greater trust in online information because many had psychiatrists who accessed online information during consultation.

While these results reflect general perceptions of trust in online information, there is little direct evidence of how trustworthy people perceive online assessment to be. As earlier mentioned, Van Ameringen et al. (2010) found roughly one in five participants completed the MACSCREEN to confirm a heath professional diagnosis, suggesting these participants viewed the program as providing a credible diagnosis. Parker et al.’s (2013) results showed a large proportion of people not only describing the MSQ program as informative, but also seemingly acting on this by undertaking follow-up activity, therefore suggesting confidence in the program.

A program’s trustworthiness could be influenced by various website-related factors. Perhaps an obvious influence is the disclaimer usually accompanying web-based feedback stating how results do not formally measure a mental disorder (Buchanan, 2003). General website characteristics may also influence trust. Klein et al. (2010) surveyed consumers of drug and alcohol related information websites and found over 70% of participants agreed that trust was enhanced when a website: provided evidence to support its claims; referenced information sources; and was clearly created by experts. Factors least frequently judged as important were if the website: was recommended by peers or family; had the HONcode seal; and is recommended by another website (Klein et al., 2010).

The low impact of third party seals (e.g. HONcode) was also observed in a qualitative study of trust in e-health websites, and was linked to peoples’ unfamiliarity with such seals (Faja & Likcani, 2005). In the same study, trust also appeared to be improved by a website’s social presence, or in other words, its ability to elicit feelings of human contact, warmth, and sociability through features such as having a picture of a therapist or personal details of others (Faja & Likcani, 2005). In another qualitative study involving focus groups and observation, general internet consumers associated a website’s credibility with: language and ease of use; the website’s source; a professional design; and official or scientific involvement (Eysenbach & Kohler, 2002). These various factors may similarly contribute to the credibility of web-based assessment programs, and subsequently influence the impact they have on respondents.

2.7 Summary of user experience.

While a web-based assessment program might be psychometrically accurate, consumer satisfaction and preference may be major determinants of whether it will be used. Underlying this is likely a range of inter-related factors such as the needs and circumstances of the user, the program's usability, and the overall experience including the short and long term repercussions of using the program. There is little empirical research examining the experience of web-based clinical diagnostic assessment programs. However, there is considerable literature regarding the overall online experience as well as computerised programs and traditional assessment approaches which could be applicable.

With greater attention to design by developers, and widespread use of the internet by consumers, it is likely that web-based clinical assessment programs can be easy to use, even for people with severe mental disorders and cognitive impairment. However, there are risks of adverse experiences. As observed with clinical interviews, the sensitive nature of clinical questions and feedback could be distressing for some, perhaps more vulnerable individuals. Also, the computerised administration format involves cognitive demands which could cause stress and fatigue. Given that web-based assessment can be conducted with relative anonymity, it seems likely it will elicit less embarrassment and greater self-disclosure of sensitive topics compared with in-person assessment as well as computerised assessment conducted in a formal setting. Together with the convenience and flexibility of use, these characteristics are likely central to the appeal and actual experience of web-based clinical assessment.

For consumers using them as a self-help tool, web-based clinical assessment could enhance self-reflection, be informative by raising awareness of mental health issues, and ultimately act as an intervention. As seen in Parker et al.'s study (2013) there is some empirical evidence suggesting web-based screening of bipolar disorder leads to help-seeking, positive lifestyle changes, treatment uptake, and reduced symptoms. However, it is uncertain whether these outcomes extend to other programs targeting different disorders and for people with different backgrounds. More research is needed to investigate how web-based programs influence people's awareness of mental health issues and behaviour (Ogles et al., 1998).

2.8 Limitations of web assessment.

While previous studies suggest web assessment can produce similar diagnostic results to a structured clinical interview, it is apparent the information collected from most web assessment programs at present is relatively limited in comparison to what can be potential

elicited in a traditional clinical interview. For example, web assessment programs, as the case with standard clinical assessment tools (e.g., paper-pencil questionnaires) often have a narrow focus and overlook important characteristics (such as personality factors, mental state) which may pertain to clinical assessment and influence the therapeutic process (Barak et al.,.2009). Also, most web assessment programs show little flexibility in being able to explore areas of importance (e.g., suicide risk) and background information. At present, these processes are only really achievable through the involvement of a clinician using direct modes of internet-assisted communication and/or a large and intensive battery of online clinical instruments.

The limits of web-based questionnaires reflects the distinction between online testing and assessment as highlighted by several researchers (Barak & Buchanan, 2004; Naglieri et al., 2004; Riva, 1998). As Naglieri et al. (2002) notes, the two terms are often used synonymously but refer to different concepts. Online testing generally refers to the standardised administration and scoring of an online instrument, where results are interpreted in reference to group norms (Naglieri et al., 2004). Accordingly, it is a nomothetic approach with a focus on the method and the materials used (Riva, 1998). In contrast, assessment refers to the idiographic process of understanding the individual in the context of a referral question, and underlying this are elements such as problem solving, hypothesis planning, and evaluation (Sloves, Docherty, & Schneider, 1979). While online testing assists clinical assessment, a formal and thorough clinical assessment comprises multiple sources of information (e.g. battery of testing, observation, background interview) assimilated with the judgment of a clinician. For the purposes of diagnostic assessment and referral, it is arguable as to whether web assessment can produce an assessment outcome comparable to a clinician-administered diagnostic assessment.

2.9 Web assessment in e-models of clinical service delivery.

The notion of delivering web-based assessment as a pathway to online treatment was raised soon after the emergence of the web (Buchanan, 1999). It continues to be promoted as a means of overcoming assessment barriers and alleviating the bottle-neck to treatment (Christensen & Hickie, 2010) and conducting research trials (Carlbring et al., 2002). As earlier mentioned, recent preliminary reports of an online system following this service delivery model supports its efficacy (Klein et al., 2011, 2012) and is consistent with the large evidence base behind online self-help resources in general (Barak et al., 2009; Griffiths, Farrer, & Christensen, 2010). Importantly, openly available and fully automated online assessment tools are proving to

be highly feasible as reflected by the large volumes of access (e.g., Parker et al., 2003; Klein et al. 2011; Van Amerigen et al., 2010).

However, there are few web-based clinical diagnostic programs readily available that have demonstrated characteristics beyond that of a basic screening tool for specific mental disorders, which therefore restricts the role and function of web assessment. Furthermore, few programs offer results without warning that they need to be clarified by a health professional. While this covers duty of care and is intended to protect consumers (e.g., in the case of invalid results and adverse experiential effects), such programs disregard the accessibility issues of those who refuse to or cannot easily access a clinician. They also overlook the fact that while assessment with a clinician may be considered ideal, it can also come with limitations (e.g., inconvenience, cost, risk of inaccuracy).

An important question arises as to whether web assessment can potentially produce sufficient diagnostic accuracy and utility to offer a suitable alternative to traditional diagnostic assessment approaches. If so, there would be greater incentive to use web assessment not just as an adjunct tool for clinicians, but also as an optional procedure in producing diagnosis and treatment recommendations in a online service delivery model without (or with minimal) clinician involvement. More specific questions include: whether web assessment should be limited to certain disorders; for whom it is more suited to; and whether web assessment requires greater control and supervision as seen in other assessment contexts (e.g. recruitment selection in organisations). Despite the benefits offered by web assessment and suggestions that it could facilitate important clinical assessment roles such as diagnosis and treatment selection (ITC, 2005), there appears to be limited application of this in both online and traditional service delivery models, and this is likely due to the lack of comprehensive diagnostic programs available and empirical evidence.

2.10 Conclusion and an opportunity for delivering online assessment.

The internet has offered new ways of delivering clinical assessment to cater for the many people going online. Web-based clinical assessment programs in particular are associated with many practical benefits for consumers and healthcare professionals while demonstrating promising psychometrics and signs of positive user experience. However, despite the many web assessment programs freely available, few known programs assess mental disorders with depth, address co-morbidity by targeting multiple disorders, and provide meaningful diagnostic and referral information. Importantly, few web-based diagnostic assessment programs provide

sufficient psychometric and user experience related evidence, which contrasts with ethical and professional guidelines recommending informed consent of online assessment use (ITC, 2005; Schulenberg & Yutrzenka, 2004)

Therefore, there was a perceived opportunity to develop and subsequently evaluate a thorough web-based clinical diagnostic assessment program and the following chapters describe this process. There were various factors to consider when deciding the type of web-based assessment program to research. Previous programs have focused on offering very brief diagnostic screening instruments or clinical measures based on existing paper-and-pencil scales. However, such programs have shown limited diagnostic breadth and depth, partly attributable to the brevity of items which results in a high level of false-positive results, and indirect or limited association with diagnostic criteria. CAT based on IRT offers promise as a potential web-based tool, and could offer a high level of precision while minimising presented items. However, CAT development is demanding while the approach is arguably too novel at present for diagnostic application.

On the other hand, web-based systems resembling existing diagnostic interview schedules appear relatively easy to develop (e.g., given there are existing schedules to model off), have closer alignment to diagnostic criteria, and have shown an acceptable level of diagnostic accuracy, at least amongst a clinic research trial population (e.g. Farvolden et al., 2002). Previous computerised programs based on diagnostic interview schedules have also shown clinical validity and been well received by consumers with mental illness, suggesting that similar web-based programs could also exhibit positive characteristics such as being easy to use and understand, as well as helpful in promoting insight. Hence, it was decided that the web-based program of focus in this research, as outlined in the following chapter, would follow the form of a structured diagnostic interview in assessing a range of disorders using self-report and text-based items corresponding to diagnostic criteria.

3.0 Chapter Three: The Online Psychological Assessment (e-PASS) Program

3.1 Overview of Chapter Four.

The following chapter outlines the development of the Online Psychological Assessment (e-PASS) and referral system of focus in subsequent studies and the general discussion. The chapter focuses on the context and constraints of the program, and includes details about e-PASS features which distinguish it from other web-based screening programs.

3.2 Introduction to the e-PASS.

The development of the e-PASS emerged in 2008 as part of a larger project awarded to Klein and Austin from the Swinburne e-Therapy Unit to develop an innovative web-based platform offering a range of online mental health resources. The platform, called Anxiety Online (AO), was conceptualised as representing a "virtual clinic" featuring an automated multi-disorder clinical diagnostic assessment program (i.e. the e-PASS) integrated with various self-help or therapist assisted treatment programs for major mental disorders, and an online eTherapist training program (Klein et al., 2011, 2012). To the knowledge of the author, this project was the first of its kind to be carried out in Australia (if not worldwide) and be publically available at the time. The project was implemented by the National eTherapy Centre (an entity associated with the Swinburne e-Therapy Unit) and funded by the Australian Government's Department of Health and Ageing, as part of an initiative to develop online mental health services to compliment more traditional mental health services (e.g. telephone, face-to-face; Klein et al., 2011, 2012).

3.3 The brief.

A primary objective for the proposal of the e-PASS was to offer an open and easily accessible web-based program that could accurately produce diagnostic results for a range of mental disorders. This included identifying the severity of diagnosed disorders and distinguishing the primary diagnosis. The e-PASS program was required to offer useful diagnostic and referral information to the respondent, with the aim of raising awareness of probable mental health issues and facilitating access to appropriate treatment based on e-PASS results.

3.4 Development of the e-PASS.

Robins and Cottler (2004) outlined seven steps for the successful development of a structured diagnostic interview schedule: 1) writing items that are consistent with diagnostic

nomenclature (e.g. DSM-IV framework); 2) ensuring target users are willing and able to respond to items; 3) selecting an acceptable format which maximises the accuracy of selecting and recording responses; 4) including a data entry process that recognises and cleans up errors; 5) creating a diagnostic scoring system that is consistent with diagnostic nomenclature; 6) developing a training program for interviewers to maximise reliability; and 7) converting the schedule for computer administration. Where relevant, these steps were closely adhered to in the development of the e-PASS and of focus in the present research.

A master document detailing the content (e.g., questions, branching rules) required to develop the e-PASS was written by Klein and the present author. This "blue print" underwent a process of extensive review and revision involving clinical psychology staff (academic and/or practicing) to ensure the face validity of the content. The finalised document was referred to by IT developers who built the e-PASS using a customised survey platform within the AO IT system. As the IT developers were not involved in the writing of the e-PASS and had limited mental health knowledge and training, the e-PASS authors liaised at length with the IT developers to clarify the requirements outlined in the master document.

The design of the e-PASS user interface took into consideration key principles of web design (see "Usability of the e-PASS" section below) which are consistent with those adopted by other websites intended for use by people with mental illness (Chinman et al., 2004; Rotondi et al., 2007; van der Krieke et al., 2012). Once produced online, the e-PASS underwent "beta testing" to ensure branching rules, diagnostic rules, and other processes were properly functioning. At a "back end" coding level, this involved staff ensuring certain combinations of response inputs would lead to appropriate diagnostic output. Higher level testing involved clinical psychology staff responding to the e-PASS in a manner consistent with certain psychiatric conditions to test if it resulted in expected diagnostic feedback.

3.5 e-PASS features.

The following section outlines specific aspects of the e-PASS and the considerations behind them.

3.5.1 Breadth of disorders.

The e-PASS assesses for 21 DSM-IV-TR disorders including (APA, 2000): the major anxiety disorders; MDD; three eating disorders; two sleep disorders; four substance dependence disorders; pathological gambling; and two somatoform disorders. Deciding on the breadth of

disorders considered several factors, such as the impact on diagnostic accuracy, the demands on the respondent, and the relevance of disorders. For example, an assessment program covering a larger range of disorders might help identify co-morbidities or inform differential diagnoses. However, a greater breath of disorders requires more items, which implies longer administration time and increased burden on respondents. This issue of response burden is pronounced when considering disorders with low prevalence rates in the population of expected users of the e-PASS. As the e-PASS was to appear on the AO website, which in its initial release focused on anxiety disorders, the e-PASS was designed to target the anxiety disorders as well as commonly associated disorders such as MDD. It was assumed that doing so would best cater to the needs of e-PASS users.

Despite their relationship and possible co-morbidity with the targeted disorders, some clinically significant disorders (e.g. bipolar disorder, schizophrenia, and personality disorders) were not thoroughly covered by the e-PASS because of their relatively complex diagnostic criteria and the difficulty in assessing them using an online self-report format. For example, the accurate diagnosis of bipolar disorder usually relies on a range of information (e.g., symptom history, collateral information) and high level of clinical judgment (Perlis, 2005) which were deemed beyond the e-PASS's scope. However, the e-PASS does include questions regarding symptoms of mania and psychosis for screening and referral purposes. Finally, the e-PASS focuses on current disorders as it is primarily concerned with presenting problems, and avoids issues of retrospective diagnosis such as erroneous recall of symptoms (Schwarz & Oyserman, 2001).

3.5.2 Diagnostic criteria.

The DSM-IV-TR classification system (APA, 2000) was chosen as the diagnostic criteria underlying the e-PASS because of its wide-spread use and acceptance across the mental health field. DSM-IV-TR criteria are also highly operational which facilitates inter-rater reliability, the communication of diagnostic information and, more specifically, allows criteria to be easily assessed using the self-report questionnaire format underlying the e-PASS. For these reasons, the DSM-IV-TR criteria have been used in many major assessment tools, including well regarded structured interview schedules. The e-PASS was designed to operate similarly to a schedule such as the MINI, in presenting items corresponding to diagnostic criteria, using branching rules based on previous responses, and having scoring rules to determine diagnostic outcome.

There are, however, drawbacks to basing the e-PASS on DSM-IV-TR criteria which should be considered. A well known practical limitation of the DSM-IV-TR system is the

considerable overlap in symptom criteria and lack of exclusionary criteria between disorders which can contribute to reliability issues and excessive diagnostic comorbidity (Pincus, Tew, & First, 2004). Regarding validity, a major criticism of the DSM-IV-TR is that its classification system arbitrarily nominates and groups symptoms to distinguish categories of disorders without sufficient theoretical or empirical justification (Dalal & Sivakumar, 2009). However, proponents of the DSM-IV claim its categorical approach of defining clinical syndromes is necessary at present, given the current lack of agreement in explanatory models of psychiatric disorders (Spitzer & First, 2005). Another criticism of the DSM-IV is that its diagnostic criteria do not translate consistently across different cultures (Widiger & Sankis, 2000).

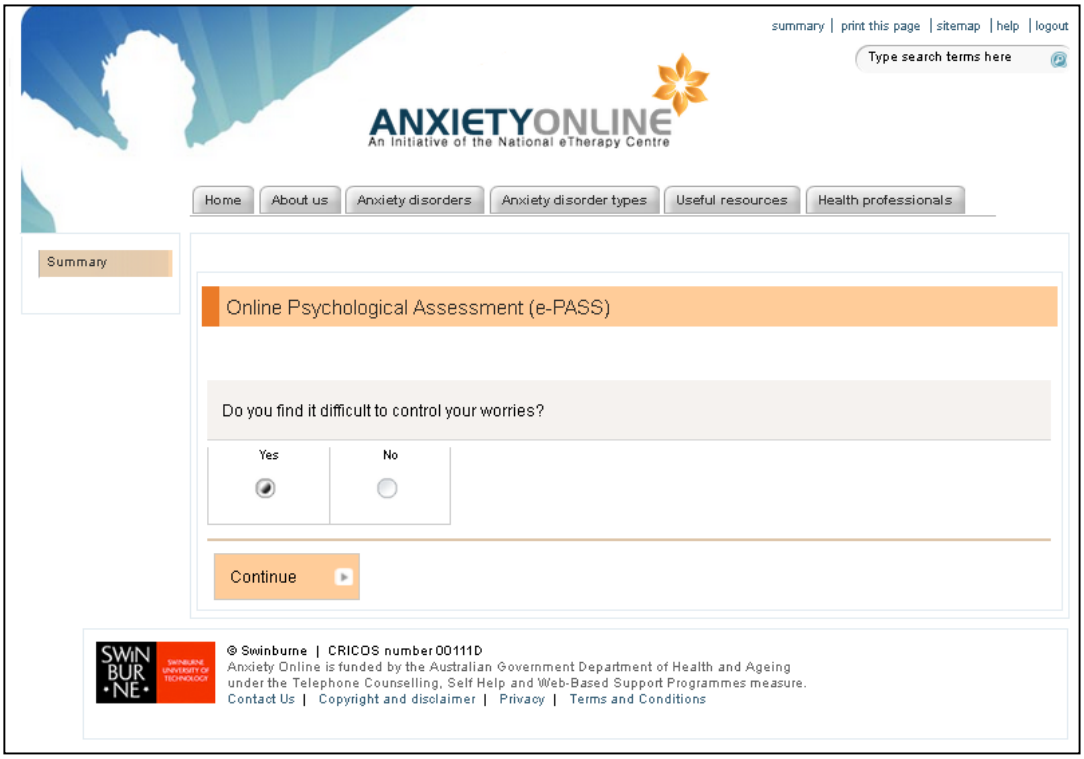
3.5.3 Items and branching rules.

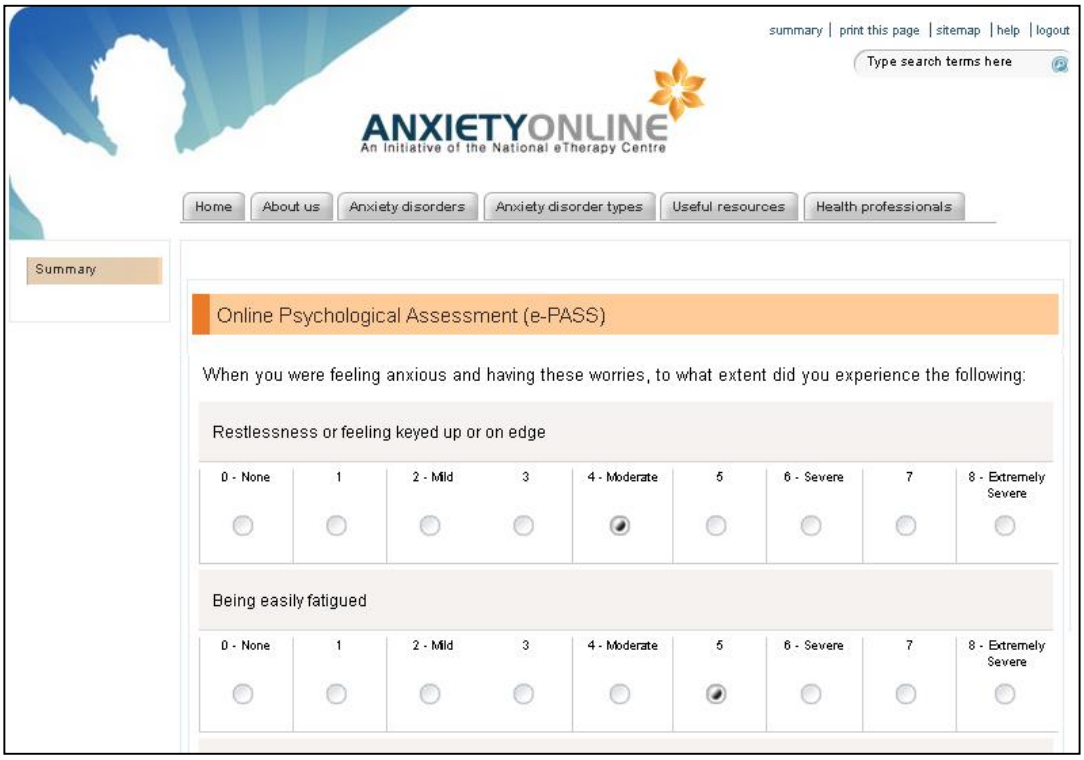
There are 540 self-rated questions in the e-PASS, consisting mainly of dichotomous (yes/no) and Likert scale questions, and a few open response questions (note: items are not included in this thesis due to university intellectual property requirements). For each disorder targeted by the e-PASS, there are questions assessing: current symptoms corresponding to DSM-IV-TR criteria; distress and impairment; period of onset; and current or previous medical conditions or substance use acting as potential casual or contributing factors. The question sets for each disorder are presented as modules. For example, the first set of questions assess for panic disorder. Within each module, questions are ordered in terms of relevance or importance to diagnosis so that questions pertaining to the most crucial features of the target disorder are presented early on and, in effect, act as screening questions.

For instance, the panic disorder module begins asking whether the respondent currently experiences distinct “periods of intense anxiety, fear, or discomfort that starts suddenly and reaches a peak within 10 minutes of beginning”. Branching (skipping) rules are used to minimise the response burden. A respondent endorsing the opening question in the panic disorder module, for example, would be presented with further questions about panic symptoms, while a person responding “no” would skip the remaining panic disorder related questions and be shown the next module (i.e., agoraphobia). As branching rules are pre-programmed, the presentation of questions appears seamless to the respondent. In effect, all respondents are presented with the same initial questions of each module, and are only presented further questions from a module if they have endorsed the initial “screening question”. Therefore, e-PASS respondents typically see a subset of the 540 e-PASS items, with item sets varying from one individual to the next

depending on symptom profile and selected responses. Figure 1 depicts two screenshots items from the GAD module of the e-PASS.

Figure 1. Two screenshots of items from the e-PASS GAD module





3.5.4 Measuring severity.

An important goal for the e-PASS was to be able to distinguish severity and recommend appropriate referral options. While there are various ways an assessment tool could determine severity, the e-PASS was designed to measure severity of a diagnosed disorder based on the overall disorder-related distress and interference reported by the respondent. If a respondent answers sufficient questions to effectively meet diagnostic criteria for a disorder, they are presented with a set of severity questions asking the participant to rate both their level of bother (or distress) and interference caused by their symptoms in "everyday living" (e.g. activities of daily living), social functioning (e.g. relationships with family and friends), and occupational functioning (i.e. at work or in studies). These severity questions are each rated on an 8-point Likert-type scale (0 = *absent*, 8 = *extremely severe*), and the same set of questions is presented at the end of each disorder related module, should an individual answer sufficient questions to reach it.

For each disorder, the scores of the severity-focused questions are averaged to determine the respondent's overall severity score for the particular disorder diagnosed. A severity threshold of 3.5 and above is used to classify a clinical disorder for most modules. A score of 3.5 generally corresponds with a "mild" rating when responded to by the user, and was deemed appropriate for balancing the sensitivity and specificity of diagnostic classification. However, it was expected that this score could require adjustment following psychometric evaluation. A clinical disorder is further categorised by severity (mild, moderate, severity) depending on where the average severity score sits between 3.5 and 8.0. A subclinical result is assigned when severity is below 3.5 or respondent endorses most but not all diagnostic criteria. The intended purpose of this subclinical label was to help people recognise and address early warning signs or vulnerability to developing a clinical disorder.

3.5.5 Distinguishing the primary from secondary diagnoses.

In the case where multiple disorders are diagnosed, the e-PASS discerns the primary diagnosis in order to help prioritise treatment recommendations. If two or more disorders have been diagnosed, the e-PASS assigns the disorder with the highest severity score as the primary disorder. When two or more disorders share the highest severity rating score, the respondent is asked to nominate the primary diagnosis which is then reflected in their final feedback. However, this process is bypassed for select combinations of disorders, where one of the disorders has been pre-determined by the e-PASS authors as having greater clinical importance and therefore requiring priority in treatment. As example of this situation is when anorexia nervosa is diagnosed

alongside other disorders and, because of its clinical importance, is deemed the primary diagnosis by the e-PASS. The decision rules around prioritising certain disorders for primary diagnosis were based on the clinical judgement and experience of the developers and associated clinical psychology staff.

3.5.6 The usability of the e-PASS.

From a usability perspective, it was important that the e-PASS be simple to use, clear, and easy to access for the respondent. All text presented in the e-PASS was written for a basic reading level. Layman's terms and descriptions were applied in the e-PASS questions and feedback, while overly technical or complex terms were avoided where possible. In addition, the developers made an effort to simplify the process of accessing and completing the e-PASS. This involved minimising the registration process, as well as the effort required by the respondent in answering the e-PASS questions. As a result, the majority of the e-PASS questions involve response selection rather text input. Other steps taken to improve the usability of the e-PASS included keeping web pages minimal and free of unnecessary content and grouping questions into pages to reduce the amount of web page scrolling required by the user.

3.5.7 Undertaking the e-PASS.

All individuals are required to create an account in order to access the e-PASS or treatment programs on AO. This process involves supplying an email address and setting a password. From there, new registered users are encouraged to undertake the e-PASS, and informed it is a pre-requisite to accessing any of the treatment programs. Those who opt to undertake the e-PASS are first presented with the terms and conditions of using the e-PASS and AO service, which highlights, amongst various points: that AO services, including the e-PASS are not appropriate for people in crisis, or requiring specialist medical treatment; that the e-PASS is undergoing formal psychometric testing; and that the e-PASS "is not intended to substitute for traditional best-practice clinical assessment by a mental health professional". If users accept this, they are presented an outline of what to expect when completing the e-PASS, then a series of preliminary questions which ask about: how the user found out about AO; demographic variables such as age, gender, country of birth; their work and study background and current status; previous and current mental health treatment; physical health issues; current quality of life; stressors and coping strategies. There are also screening questions for thought and perceptual disturbances, and suicidal ideation. Finally, the Kessler 6 questionnaire (Kessler et al., 2003) is presented as part of the pre-assessment.

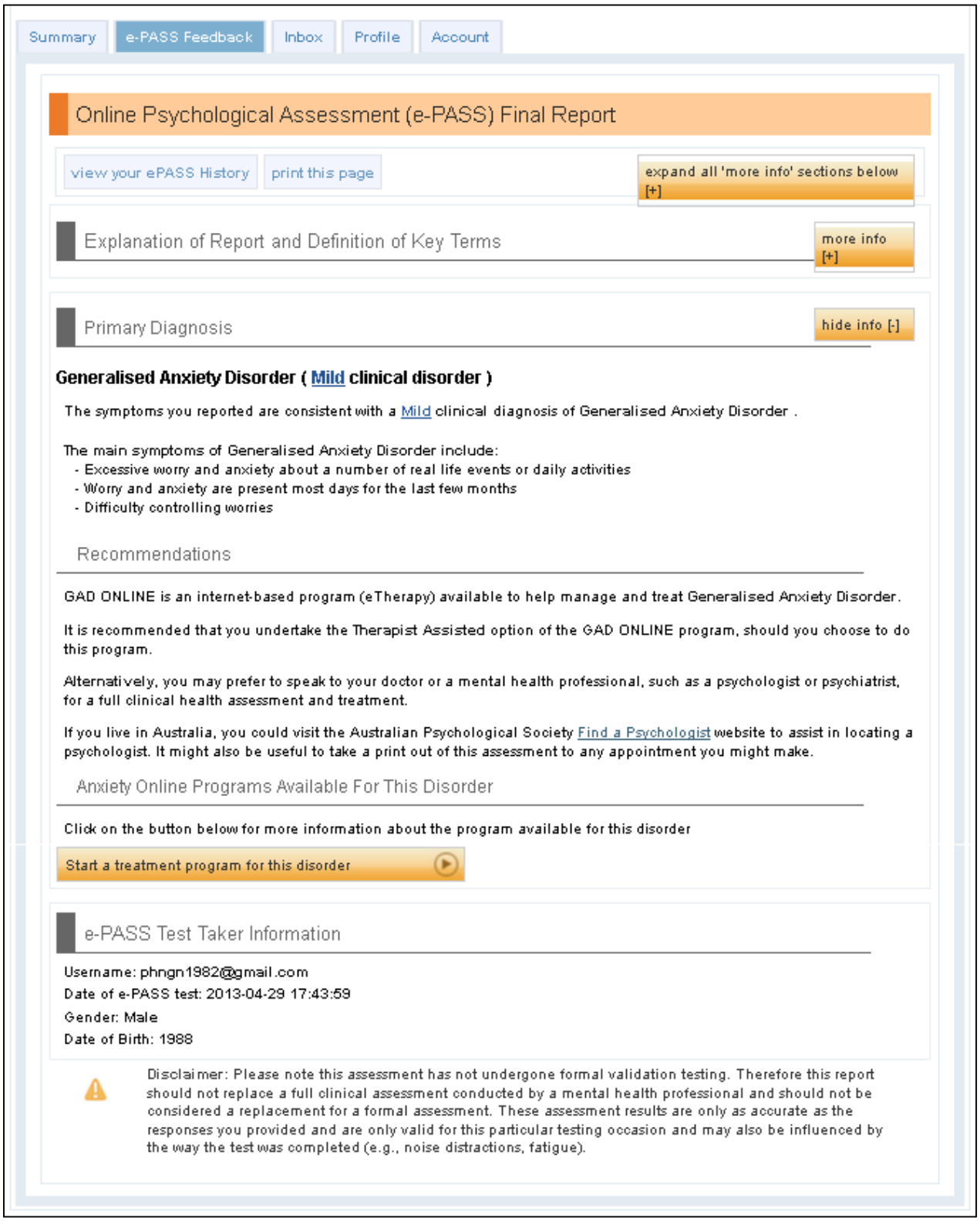
The e-PASS is stated as taking 10-60 minutes to complete and varies in length according to the number of questions presented (which in turn depends on what symptom groups are relevant to the respondent) and how respondents approach answering questions (e.g., quickly, without much consideration, or slowly and carefully). Respondents can take breaks or exit the e-PASS (intentionally or unintentionally) and resume from their last presented question, but must complete the e-PASS within 24 hours to produce a result; otherwise, the program resets to the initial assessment item. After completing all relevant questions, respondents are presented a feedback page.

3.5.8 e-PASS feedback.

The provision of assessment feedback is generally recommended to help educate respondents about their symptom profile and act as a guide to treatment (Dozois & Dobson, 2010). To address this, the e-PASS concludes with a feedback page which respondents can print out or refer back to in future via their AO user account (see Figure 2 for example). At the top of the feedback page is a general explanation of the report including definitions of key terms used. For example, the respondent is given a brief explanation distinguishing “primary” and “secondary” diagnoses, and the difference between a “clinical” and “subclinical” disorder. Following this is a description of the primary and, if present, secondary diagnoses. For each diagnosis listed, the name of the disorder, its assigned severity category (i.e., subclinical, mild, moderate, severe) and the main symptoms of the disorder are outlined. Finally, referral recommendations are provided which cater to the diagnosed disorder(s) and its severity.

Generally, respondents diagnosed by the e-PASS with a disorder targeted by one of the AO treatment programs are recommended to undertake the self-help version of that particular program if the disorder is of *subclinical*; or the therapist-assisted version of the program if the disorder is of *mild*, *moderate* or *severe* category. Other online treatment programs outside of AO are recommended where relevant. For example, prior to the release of Depression Online (a depression focused treatment program on AO) respondents given an e-PASS subclinical diagnosis of MDD are recommended the MoodGym online self-help program (Bennett et al., 2010) for depressive symptoms. Respondents are recommended to see a mental health professional when: an online treatment program is not available; when a medical condition or substance use has been reported as potentially impacting on symptoms; disorders are of high severity, or risks such as suicidality or psychotic symptoms are detected by the e-PASS.

Figure 2. Example of e-PASS feedback report. This report corresponds with a response profile (artificially generated) representing mild GAD. The recommendation of GAD Online refers to the online treatment program for GAD at Anxiety Online. The "Explanation of Report..." section has been minimised to fit the report on this page.



3.6 Preliminary e-PASS data.

Shortly after the e-PASS was created and prior to its launch, a pilot study was conducted by the author with the aim of ensuring the e-PASS could be adequately performed by people with mental health issues before it was to be widely used in the following research studies (unpublished; Nguyen, 2013). Participants (7 men and 4 women, average age of 30.1 years) were clients at a university-based psychology clinic, predominantly undergoing group therapy for social phobia or OCD. Participants underwent an in-person clinical assessment involving the MINI-Plus conducted by clinic staff (provisional or fully registered psychologists), and subsequently completed the e-PASS in a time and setting of their choosing. Participants were reimbursed with a book voucher.

With MINI-Plus results as the validity criterion, classification statistics were calculated for the e-PASS diagnostic result of OCD, GAD, panic disorder, social phobia, and MDD. Sensitivity varied from .33 to 1.00 while specificity was generally higher, and ranged from .83 to 1.00. Positive predictive values were particular strong for social phobia and OCD (both 1.00) while the negative predictive value was at least .71 across different disorders. Participants were also asked to report any major issues with e-PASS completion. Only one participant reported an issue in which the e-PASS "timed out" and restarted, and this was attributed to an extended break taken during e-PASS completion. Although these results were limited by the small sample, they suggested the e-PASS was functioning adequately to warrant a more comprehensive evaluation.

There is also preliminary data comparing e-PASS results against community-based sources. As part of a larger study evaluating the outcomes of clients completing AO treatment programs, Klein et al. (2011) identified 64 participants who had, at post-treatment, sought confirmation of their initial e-PASS results with community sources, including psychologists (n = 33), medical doctors (n = 16), and counselors (n = 4). Almost all participants (95%, n = 61/64) reported agreement between their e-PASS result and these community sources. Disagreement (n = 3) was noted between the e-PASS and a medical doctor, a website, and a friend. As Klein et al. (2011) acknowledged, these results provided a crude indicator of the e-PASS's validity, given participants' reports could not be verified and the selective sample involved.

3.7 Aims of the present research.

The decision to employ a particular assessment program should be made with regard to its various advantages and disadvantages as well as its suitability for the needs of the situation (Noyes & Garland, 2008). While intended to offer a range of benefits (e.g., accurate multi-disorder

diagnostic results, convenient, ease of use), the e-PASS has limited evidence of actual performance. This represents the lack of empirical evidence of web-based clinical assessment programs in general, but particularly of applications that are more comprehensive in identifying co-morbidity, addressing diagnostic criteria, and providing diagnostic feedback to the consumer.

Therefore, the main aim of this research was to evaluate the e-PASS in real world conditions and focus on two important areas, namely the psychometric properties and the user experience of the e-PASS. When evaluating a web-based clinical diagnostic program, as with any assessment instrument, the primary criterion is traditionally its psychometric properties. A program lacking in validity and reliability can have significant implications such as misdiagnosis, inappropriate treatment recommendations, and longer term mismanagement of mental disorders. Hence Study One investigates the validity as well as the test-retest reliability of an e-PASS diagnostic result.

Web-based programs such as the e-PASS appear to be used by a large number of people in a range of situations and with various needs. One likely scenario is of mental health consumers directly accessing the e-PASS on their own, with the aim of receiving meaningful diagnostic results. It is unclear as to how different people perceive, perform, and respond to the e-PASS. However, their experience forms an integral part of the program's clinical utility. Therefore, the user experience of the e-PASS was explored with a predominantly quantitative approach in Study Two, and a qualitative approach in Study Three. The resulting mixed-methodology is considered advantageous in eliciting depth and breadth of experience at both an individual and group perspective (Johnson, Onwuegbuzie, & Turner, 2007).

3.7.1 Stages of development and evaluation.

The Stage Model of Behavioural Therapies Research outlines three progressive stages of development and evaluation of behavioural interventions which are considered ideal for web-based intervention as they facilitate innovation and dissemination (Danaher & Seeley, 2009). Stage one involves a formative evaluation including an assessment of the problem and needs of the target population, evaluation planning, communication strategies, and an assessment of specific system requirements (e.g., features, user interface). Stage two focuses on evaluating whether the intervention achieves its intended outcome under ideal conditions and usually follows an experimental design. The third and final stage examines whether the outcomes of stage two can be carried over to more real-world contexts, and addresses issues such as whether (Danaher & Seeley, 2009): the intervention can be implemented on a wider scale; intervention effects can

be generalised across different population groups; the intervention is cost-effective; and the intervention is socially valid or accepted.

For this research, given the current delivery of the e-PASS, there was an emphasis on stage three of the evaluation process, with a focus on evaluating the effectiveness of the e-PASS as a diagnostic instrument when used in real-world settings. The assessment of the problem, needs, system requirements etc. pertaining to stage one occurred in the process of developing the e-PASS. As earlier outlined in this chapter, this largely involved the review of relevant literature and consultation with various experts; usability testing by developers to ensure the technical and functional adequacy of the e-PASS; and a pilot study of validity and usability with a small sample of clients from individual and group therapy at a psychology clinic.

Due to constraints in time and other resources, the evaluation of the e-PASS in controlled conditions (stage two) was not feasible for this research. Furthermore, it was argued that evaluating the e-PASS under controlled conditions could diminish some of the effects and phenomena (e.g., relative anonymity and how this might relate to diagnostic accuracy of e-PASS) of interest in this research. Therefore, it was decided to conduct stage three and evaluate the e-PASS by recruiting actual e-PASS participants (i.e., rather than recruit from a university or clinical trial) just prior to completing the e-PASS, and allowing participants to complete the e-PASS in an environment and manner of their choosing. On the one hand, this design potentially conflates factors that may contribute to the accuracy and experience of the e-PASS.

However, there are several benefits of this methodology. Although a form of convenience sampling, recruited participants are likely more representative of the population who use the e-PASS. They also presumably present a wider distribution of socio-demographic and mental health related characteristics than would otherwise arise if recruiting participants from more convenient populations (e.g., university students, clinic clients; Reips, 2002). In addition, participants display a higher degree of voluntariness because they have an intrinsic need to perform the e-PASS and have fewer external constraints on commencing and continuing participation (Reips, 2002). Finally, participants complete the e-PASS in real-world circumstances, therefore enhancing the ecologically validity of the results and findings (Reip, 2002) which was deemed a priority of this research. Together, the inclusion of more heterogeneous participants and settings for evaluating the e-PASS reflects the main differences between stage three and two of the Stages Model (Danaher & Seeley, 2009).

3.7.2 Additional aims and objectives.

The results and findings of this research will also help clarify the potential role and value of web-based clinical diagnostic assessment programs in addressing the needs of people with mental health issues. This research will also consider how a web-based program such as the e-PASS could fit into the wider context of clinical assessment and treatment. As psychometric and experiential issues will be identified, the final objective of this research will be to propose recommendations for enhancing the e-PASS. These points will be referred to throughout the three studies, though will be summarised and discussed in the General Discussion and Conclusion.

4.0 Chapter Four: Evaluating the Validity and Reliability of the e-PASS (Study One)

4.1 Introduction.

The e-PASS is a web-based clinical assessment tool for the diagnosis of mental disorders. In many ways, it resembles the structure and content of existing or previously reported computerised and online programs (e.g. CIDI-Auto, WB-DAT). However, the e-PASS is relatively distinct in terms of its particular appearance, context (e.g., Australian developed and integrated with treatment programs on the Anxiety Online [AO] website), availability (e.g., free, 24/7), and assessment focus (e.g., subclinical/clinical diagnoses of many anxiety, mood, eating, and substance disorders). While previous psychometric literature in relation to online clinical assessment programs exist, they may not necessary reflect the properties of the e-PASS.

The ITC (2005) guidelines on internet testing states test developers must pay due attention to the provision of accurate feedback (Bartram, 2006). As the e-PASS is currently available for public use, has a wide reach, and informs treatment recommendations, there is a strong need to ensure the e-PASS performs with an acceptable level of diagnostic accuracy. There is some evidence of validity seen in a previous study where the majority of a small sample who compared their e-PASS results with an external source (e.g., health professional) reported agreement (Klein et al., 2011, 2012). However, more comprehensive evidence of its psychometrics is needed to clarify the e-PASS's suitability as a clinical diagnostic tool.

This study aimed to evaluate important psychometric properties of the e-PASS, namely the criterion validity and test-retest reliability of an e-PASS clinical diagnosis. This study builds on the preliminary psychometric data gained through pilot testing and community agreement rates observed in AO users. Unlike previous evaluations of web-based clinical assessment programs, participants were recruited directly after accessing the e-PASS and were therefore representative of those who actually use the e-PASS. The validity of an e-PASS diagnosis was measured against several criteria. As commonly employed, a structured clinical interview (CI) was the main criterion and involved two well-regarded interview schedules: the MINI-Plus (MINI) and the ADIS-IV. The MINI, as reported in other studies, could have been solely used as validity criteria for the majority of targeted disorders. However, the ADIS-IV was added as it is considered a gold-standard interview schedule for anxiety disorders, and includes questions (e.g., regarding differential diagnosis, function of symptoms, severity of impairment/distress) that exceed the scope and detail elicited in standard schedules (Grisham, Brown, & Campbell, 2003).

The e-PASS was also examined against widely used standardised measures of symptoms (e.g., Y-BOCS-SR, CES-D), impairment, and disability (e.g., K-6), which are also considered important criteria for diagnostic validity (Robins, 1994). While self-administered questionnaires are not considered a diagnostic gold-standard, several have strong diagnostic characteristics and are commonly used for assessment purposes (e.g., screening, treatment/outcome measure) in research and practice. Of interest was whether an e-PASS diagnosis and severity for a particular disorder (e.g., major depressive disorder [MDD]) associated with a higher and diagnostically meaningful score on a corresponding questionnaire (e.g., CES-D which measures for depression severity). An additional area of interest was how diagnostically accurate the e-PASS was compared with the questionnaires, using the structured clinical interviews as the gold-standard criteria.

Given the diagnostic breadth of the e-PASS, the psychometric investigation targeted a number of disorders, some of which have not been reported in the web-based clinical screening literature. For example, no known studies have looked at whether bulimia nervosa (bulimia) and body dysmorphic disorder (BDD) can be effectively identified with a web-based program. More generally, there have been no known studies which have evaluated both the criterion validity and test-retest reliability of a web-based program for the screening of a range of mental disorders and involving a sample of actual users of the program. However, based on the face validity of the e-PASS and the promising findings of other web-based programs such as the WB-DAT (e.g., Farvolden et al., 2002), WSQ (Donker et al., 2008), and the ISP-D (Lin et al., 2009), the e-PASS was expected to display adequate criterion validity and test-retest reliability.

4.2 Methodology.

4.2.1 Considerations.

Guidelines have been proposed by researchers to promote good practice of reporting studies of diagnostic accuracy. Referred to as the "STAndards for the Reporting of Diagnostic accuracy studies" (STARD), the STARD statement (see Appendix B) is a checklist of 25 items to address. These include the need to describe (see www.stard-statement.org): the study population and recruitment procedure; how data is collected; the reference standard (e.g., validity criteria) and its rationale; technical specifications of materials and method used; the number and training of persons involved in obtaining results; methods of calculating statistical results; reporting estimates of diagnostic accuracy and measures of uncertainty; and discuss the clinical applicability of the study findings. Over 200 journals refer to the STARD statement when

instructing research authors, and these guidelines have been adhered to by numerous studies (see Meyer, 2002). Hence, where applicable, STARD guidelines were also followed in the reporting of this study. Furthermore, the research reported as follows (and in the following studies of this thesis) received approval from Swinburne University Human Research Ethics Committee (2008/143; see Appendix A).

4.2.2 Recruitment and participants.

Although the e-PASS is open to international users, study eligibility was limited to individuals residing in Australia to allow adequate duty of care and facilitate research activities (e.g. phone interviewing). The study welcomed any Australian adult aged 18 and over with any degree of clinical symptoms related to the disorders assessed by the e-PASS. Those who considered themselves as highly suicidal or distressed were initially encouraged via the Information sheet (see Appendix C) to defer participating in the study. However, they were allowed to continue and were prompted to seek appropriate help when identified during the study as having high risk factors (e.g., assessed as highly suicidal during CI, or screened during the e-PASS as having psychotic symptoms).

All recruitment occurred through the AO website (www.anxietyonline.org.au), a university run and government-funded public online mental health service offering a range of mental health resources (including the e-PASS). Website users who clicked a link to commence the e-PASS were presented a brief advertisement (see Appendix D) inviting them to voluntarily participate in this study. This recruitment method was expected to attract individuals varying in clinical symptom severity and motivation for undertaking the e-PASS (e.g., to be formally assessed, or out of curiosity, or on behalf of others). However, it was assumed that these individuals would have at the least completed the e-PASS, regardless of their participation in this study, thus enhancing the ecological validity of the eventual data collected.

Individuals who declined the recruitment offer went onto commence the e-PASS independently of this research. Individuals who selected for further details were presented with the information sheet (Appendix C) which included an outline of the e-PASS, and the aim, inclusion criteria, and potential participation activities of the study. The information sheet stated that Australian adults aged 18 years or older were eligible to participate in the study; while those who were highly suicidal or distressed were recommended to defer participating, though could continue if they desired. Individuals eligible and interested in the study proceeded to an online consent form (Appendix C). Participants consented by supplying their name, contact details,

availability for a phone interview, and their general practitioner's (GP) contact details, to be used in case of emergency.

A total of 645 participants aged between 18 to 77 years consented to undertake the overall study, representing 7.1% of the total number of people (*N* = 9,085) who publically accessed the e-PASS during the 18 month recruitment period of November 2009 to June 2011. Twenty nine individuals were excluded and had their data removed for either residing outside of Australia or having incomplete e-PASS data. The remaining 616 participants were aged between 18 to 77 years (*M* = 38.1, *SD* = 12.7). Participant socio-demographics are presented in Table 2 while Table 3 depicts the breakdown of overall participants who went on to complete the various participation activities.

Table 2

Frequency and Percentage of Sociodemographic Characteristics of the Total Sample

	Categories	Frequency	Percentage (%)
Age (years)	18-25	88	14.3
	25-35	207	33.6
	35-45	158	25.7
	45-55	92	15.0
	>55	70	11.4
Gender	Male	174	28.3
	Female	442	71.7
Employment status	Employed full-time	237	38.5
	Employed part –time/casual	169	27.5
	Home duties	39	6.3
	Disability Support	35	5.7
	Unemployed	54	8.8
	Retired	20	3.2
	Other	62	10.0
Residential setting	Metropolitan	389	63.2
	Regional	151	24.5
	Rural	63	10.3
	Remote	12	2.0
Highest level of schooling	None	10	1.7
	Completed primary school	3	0.5
	Secondary school up to Year 10	92	14.9
	Year 11 Secondary	41	6.7
	Year 12 Secondary	469	76.2
Highest level of post-schooling	None	211	34.3
	Apprenticeship / Trade Certificate	26	4.3
	Diploma	57	9,2
	Currently completing undergraduate degree	82	13.3
	Completed undergraduate degree	126	20.5
	Post graduate degree	108	17.5
	Other	6	0.9

Note. N = 616

Table 3

Number and Proportion of Participants in the Different Stages of the Study

	No. of participants	% of total included participants
Participants consenting to the study	645	-
Participants included in the study	616	100.0
Participant who completed:		
e-PASS	616	100.0
Clinical interview	158	25.6
Online questionnaire	173	28.1
Repeated the e-PASS	39	6.3

4.2.3 Materials.

The e-PASS.

As previously described in Chapter Three, the e-PASS is an online clinical assessment schedule publicly available on www.anxietyonline.org.au. The e-PASS assesses for 21 DSM-IV disorders and provides feedback regarding diagnosis, severity (four categories of increasing severity: subclinical, mild clinical, moderate clinical, and severe clinical disorder), and referral recommendations. The e-PASS consists of 540 items, though screening items and branching rules minimise the presentation of items deemed irrelevant to the respondent and their results. On average, the e-PASS takes 25 minutes to complete.

Materials used to measure the validity of the e-PASS.

Table 4 summarises the disorders of focus in the validation study and corresponding interview schedules and self-report measures used to validate the e-PASS.

Table 4

Interview Schedules and Self-Report Measures Used to Validate the e-PASS

e-PASS Disorders	Interview schedules	Self-report measures
Agoraphobia	ADIS-IV	Fear Questionnaire
Generalised anxiety disorder (GAD)	ADIS-IV	Generalised Anxiety Disorder – 7
Obsessive-compulsive disorder (OCD)	ADIS-IV	Y-BOCS – Self Report
Panic disorder	ADIS-IV	Panic Disorder Severity Scale – Self Report (PDSS-SR)
Post-traumatic stress disorder (PTSD)	ADIS-IV	Impact of Events – Revised
Social phobia	ADIS-IV	Social Phobia Inventory – SPIN
Specific phobia	ADIS-IV	Fear Questionnaire (FQ)
Anorexia nervosa (anorexia)	MINI-Plus	SCOFF; Eating disorder Screen for Primary care (ESP)
Bulimia nervosa (bulimia)	MINI-Plus	SCOFF; Eating disorder Screen for Primary care (ESP)
Major depressive disorder (MDD)	MINI-Plus	CES-D; BDI-II
Cannabis dependence	MINI-Plus	-
Stimulants dependence	MINI-Plus	-
Opioid dependence	MINI-Plus	-
Sedative dependence	MINI-Plus	-
Alcohol dependence	MINI-Plus	AUDIT
Somatisation disorder	MINI-Plus	-
Body dysmorphic disorder (BDD)	MINI-Plus	Body Dysmorphic Disorder Questionnaire (BDDQ)
Binge eating disorder (BED)	Eating Disorders Examination–Binge Eating module	SCOFF; Eating disorder Screen for Primary care (ESP)
Pathological gambling	Diagnostic Interview for Pathological Gambling	South Oaks Gambling Screen
Primary insomnia (insomnia)	The Insomnia Severity Index	-
Primary hypersomnia	The Insomnia Severity Index	-

Interview schedules used in the clinical interview.

The following outlines the interview schedules used in the telephone CI.

The Mini-International Neuropsychiatric Interview – Plus (MINI-Plus).

The MINI-Plus (English version 5.0.0) is a brief structured diagnostic interview schedule (Sheehan et al., 1998). It assesses a range of major Axis I disorders consisting of questions based on DSM-IV diagnostic criteria. The MINI-Plus has shown good psychometric properties compared with well validated measures such as the SCID-IV and CIDI (Sheehan et al., 1998). Compared with these other schedules, the MINI-Plus is argued to be a good diagnostic instrument for research purposes as it is shorter to administer while retaining acceptably high levels of validity and reliability (Sheehan et al., 1998). Variants of the MINI have been used in numerous studies as a diagnostic validity criterion (e.g., Bunevicius et al., 2007; Cuijpers et al., 2008; Haringsma, 2004; Jardi et al., 2006; Lin et al. 2007). Corresponding with the areas of focus in this study, the following MINI-Plus modules were administered in the CI: major depressive episode; suicidality; (hypo)manic episode; panic disorder, agoraphobia; social phobia; specific phobia; PTSD; alcohol dependence; cannabis dependence; stimulant dependence; sedative dependence; opioid dependence; anorexia; bulimia; GAD; somatisation disorder, and; BDD.

Anxiety Disorder Interview Schedule for DSM-IV (ADIS-IV).

The ADIS-IV is a semi-structured interview schedule primarily designed to assess for current anxiety disorders (Brown, Di Nardo, & Barlow, 1994). Additional modules assessing mood, somatoform, and substance use disorders are included to address the high co-morbidity and overlap in symptomatology of these and anxiety disorders. The following ADIS-IV disorder modules were used for this study: panic disorder, agoraphobia; social phobia; GAD; specific phobia, and; PTSD. The ADIS-IV schedule consists of questions which reflect the DSM-IV criteria as well as additional questions to assist with differential diagnosis and functional analysis of the anxiety disorders (Hilsenroth, Segal, & Hersen, 2004).

Regarding its psychometric properties, the ADIS-IV has demonstrated reliability considered good to excellent for all diagnostic categories, except dysthymic disorder (Hilsenroth et al. 2004). There are no published studies on the validity of the ADIS-IV. However, the ADIS-R, a predecessor of the ADIS-IV, has been shown to align well with other diagnostic measures, particularly when diagnosing for panic disorder (e.g., Rapee, Brown, Antony, & Barlow, 1992). The ADIS-IV is argued to provide more in-depth clinical and diagnostic information compared to

popular instruments such as the SCID-IV, particularly regarding anxiety disorders (Hilsenroth et al., 2004). However, it is important to note the ADIS-IV takes both a categorical and dimensional approach to diagnosis. This differs from the MINI-Plus (and other schedules) which more closely follows the DSM-IV in conceptualising disorders with a purely categorical approach (Hilsenroth et al., 2004). For example, in assessing for severity and frequency of key symptoms, the ADIS-IV uses Likert-scaled questions while the MINI-Plus employs closed-ended (i.e., yes/no) questions. Therefore, a potential strength of the ADIS-IV's dimensional approach is its sensitivity to severity thus allowing more specific comparison with the scaled severity outcomes of the e-PASS.

The Eating Disorders Examination (EDE) interview – V. 16.0.

The EDE (Fairburn & Cooper, 1993) is a semi-structured interview schedule which assesses DSM-IV eating disorders including anorexia, bulimia, and binge eating disorder (BED). The psychometric properties of the EDE have been examined using community and clinical samples. The EDE has demonstrated good internal consistency coefficients (.76 to .90), test-retest reliability (.83 to .97), and inter-rater reliability kappas (.83 to .99; e.g. Cooper, Cooper, & Fairburn, 1989; Fairburn & Cooper, 1993; Rizvi, Pterson, Crow, & Agras, 1999). The BED module of the EDE was used for this study.

Diagnostic Interview for Pathological Gambling Severity (DIGS).

The DIGS (Winters, Specker & Stinchfield, 1997) consists of various items regarding areas such as demographics, gambling involvement, treatment history, onset of gambling, gambling frequency, medical status, family and social functioning. In addition, there are 21 items relating to DSM-IV diagnostic criteria which were used for this study. There is evidence that the DIGS presents good reliability and validity in identifying pathological gambling (e.g., Lesieur & Blume, 1987).

The Insomnia Severity Index (ISI)

The ISI consists of seven items and is a brief screening measure of insomnia commonly used as an outcome measure in treatment research (Bastien, Vallières, & Morin, 2001). The ISI focuses on subjective difficulties getting to sleep, staying asleep, and waking prematurely in the previous two weeks. The ISI has been found to be a reliable and valid instrument to measure subjective insomnia severity in both a population and clinical sample (e.g., Bastien et al., 2001; Morin, Belleville, Bélanger, & Ivers, 2011).

Self-report measures used in the Online Questionnaire.

The following self-report measures were presented in the Online Questionnaire.

Fear Questionnaire (FQ)

Developed by Marks and Matthews (1979), the FQ is a one page self-rating form which produces four scores: *level of avoidance* score referring to a specific phobia; a *total phobia* score reflecting extent of avoidance for 15 phobias; a rating of *associated anxiety and depression*; and a *global phobia* rating representing distress and avoidance (Marks & Matthews, 1979). There are also three subscales: agoraphobia subscale, blood injury phobia subscale, and a social phobia subscale. The FQ is widely used and has demonstrated good psychometric properties, particularly for diagnosing agoraphobia, social phobia, and blood/injury phobia (Mavissakalian, 1986). For this study, four items relating to fear of animals, natural events (e.g. lightning), body fluids, and materials (e.g. detergent) were added to the 15 items used to calculate the *total phobia* score, which changed the scoring range from 0-120 to 0-152. The total phobia score was used in comparison with the e-PASS diagnosis of specific phobia. The social phobia FQ subscale was scored between 0-40 and compared against the e-PASS social phobia diagnosis. Higher FQ scores reflect higher levels of phobia. Standardised cut-off scores could not be found for the FQ.

Generalised Anxiety Disorder Screener (GAD-7)

The GAD-7 was designed to be a brief screener of GAD. It comprises seven items (four point rating scale) referring to anxiety symptoms in the previous two weeks, and an additional item measuring the impact of symptoms on functioning (Spitzer, Kroenke, Williams, & Löwe, 2006). Scores range from 0 to 21, with four categories: normal (0-4); mild (5-9); moderate (10-14); and severe (15-21). An optimal sensitivity (.89) and specificity (.82) for detecting GAD has been found using a cut-off score of 10 (Spitzer et al., 2006). Generally, the GAD-7 has demonstrated good reliability, in addition to construct, criterion, factorial, and procedural validity (Spitzer et al., 2006). The GAD-7 has also shown to be a reliable and valid measure of anxiety in the general population (Löwe et al., 2008).

Yale-Brown Obsessive Compulsive Scale – Self Report (Y-BOCS-SR)

The Y-BOCS–SR is a measure of OCD symptom severity (Steketee, Frost, & Bogart, 1996) and is often used to facilitate diagnostic assessment of OCD. It is a self-administered alternative to the widely used clinician-administered Y-BOCS (Goodman et al., 1989). The Y-

BOCS-SR covers 58 obsessions and compulsions, which respondents indicate whether or not they experience. The respondent is then required to identify their three main obsessions and compulsions. In relation to these, five questions about time spent, interference, distress, resistance, and control, are asked on a 5-point Likert scale (0 = “none”, 4 = “extreme”). The Y-BOCS-SR has shown good internal consistency in both clinical and non-clinical samples; good test-retest reliability in a non-clinical sample; and strong convergent reliability as shown by highly agreeing with the clinician-administered Y-BOCS (Steketee et al. 1996). A cut off score of 16 has been found to produce excellent sensitivity, ranging from .94 to 1.00 (Steketee et al., 1996).

Panic Disorder Severity Scale – Self Report (PDSS-SR)

The PDSS-SR is a self-administered version of the interviewer-administered Panic Disorder Severity Scale with minor wording modifications to the original (Shear et al., 2001). The PDSS-SR consists of seven items based on DSM-IV panic disorder criteria, which respondents’ rate using a 5-point Likert scale. The scores of these items are averaged and totaled to produce a composite score, ranging from 0 to 28. Shear et al. (2001) found that a cut-off score of eight detected panic disorder with .83 sensitivity and .64 specificity. The PDSS-SR has shown good test-retest reliability and intra-class reliability with the interviewer-administered PDSS (Houck, Spiegel, Shear, & Rucci, 2002; Lee, Kim, & Yu, 2009).

Impact of Events – Revised (IES-R)

The IES-R is 22 item self-report measure of subjective distress related to traumatic events (Weiss & Marmar, 1996). It is a revision of the IES, which has been a widely used measure of trauma impact. The IES-R requires respondents to identify a traumatic event and rate how it has affected them in the past seven days, on a 5-point scale (0 = "not at all", 4 = "extremely"). The IES-R is scored from 0 to 75, with higher scores indicating more severe PTSD characteristics. IES-R items are related to 14 out of the 17 PTSD symptoms listed in DSM-IV criteria and cover the key areas of intrusions, avoidance, and hyper-arousal. While the IES-R was not designed to diagnosis PTSD, cut-off scores have been suggested which provide a provisional diagnosis of PTSD for specific population groups (Weiss & Marmar, 1996). Craemer, Bell, and Failla (2003) suggested a cut-off score of 33 and above as a result of their study with Vietnam veterans, whilst others have suggested 22-24 when used with individuals presenting substance dependence (Rash, Coffey, Baschnagel, Drobes, & Saladin, 2008) or victims of natural disaster (Asukai et al. 2002). The IES-R has shown good correlation with other recognised measures of

PTSD, including the Clinician-Administered PTSD Scale and PTSD Symptom Scale-Self Report version (Foa, Riggs, Dancu, & Rothbaum, 1993).

Social Phobia Inventory (SPIN)

The SPIN is a brief self-report screening instrument for measuring social phobia severity (Connor et al., 2000). It contains 17 items referring to social phobia symptoms (e.g. fear, avoidance, physiological) which respondents rate on a five point Likert-scale (1 = "not at all" to 5 = "extremely"). The SPIN has demonstrated excellent internal consistency and good test-retest reliability, as well as good convergent and discriminant validity (Antony et al., 2006; Connor et al., 2000). Preliminary research has found that a cut-off score of 19 is useful for distinguishing between those with social phobia and non-anxious control participants with 79% accuracy (Connor et al., 2000).

Centre for Epidemiological Studies Depression (CES-D) scale

The CES-D is a 20 item self-report measure of current depression symptoms focusing, in particular, on affect and mood (Radloff, 1977). It is widely used and one of the most popular self-administered measures of depression. Respondents rate how often they have felt in relation to item content using a 4 point Likert scale (0 = “rarely or none of the time”, 3 = “most or all of the time”). Reversal items are included to address response bias. CES-D scores are totalled (ranging 0 to 60) with higher scores indicating more severe depression. The CES-D has demonstrated good reliability and validity across a variety of populations (see Rush, First, & Blacker, 2008). CES-D scores of 16 to 26 are suggested to indicate mild depression (Ensel, 1986), while scores of 27 or more have been found to be more useful for screening MDD in medical patients (Geisser, Roth, & Robinson, 1997; Zich, Attkisson, & Greenfield, 1990).

Body Dysmorphic Disorder Questionnaire (BDDQ)

The BDDQ is a brief self-report screening measure of BDD consisting of nine items referring to DSM-IV criteria. Specifically, items screen for excessive preoccupation with an imagined defect in appearance, not better accounted for by another mental disorder (e.g. anorexia), and; clinically significant distress or impairment in functioning caused by the preoccupation (APA, 2000). All items are dichotomous (yes/no) except for the final item, which asks the respondent to list the average hours spent per day preoccupied with their defect (Phillips, 2005).The BDDQ has displayed good sensitivity (1.00) and specificity (.89) in a psychiatric setting (Phillips, Atala, & Pope, 1995).

Alcohol Use Disorders Identification Test (AUDIT)

The AUDIT is a brief screening measure of recent alcohol use developed by the WHO (Saunders et al., 1993). The AUDIT consists of 10 items covering three domains (hazardous alcohol use, dependence symptoms, and harmful alcohol use) consistent with the ICD-10 definition of alcohol dependence (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001). It was developed using a multi-national sample and has been extensively evaluated, demonstrating very good psychometric properties (Allen, Litten, Fertig, & Babor, 1997; Babor et al., 2001). The AUDIT can be administered as an oral interview or self-report questionnaire, and is deemed suitable for computer-administration. An AUDIT score of 20 or above indicates alcohol dependence and the need for follow-up diagnostic evaluation and treatment (Babor et al., 2001). In this study, participants responding positively to an alcohol use screening question (i.e. "do you drink alcohol at least once a month?) were then presented the AUDIT.

SCOFF

The SCOFF is a brief screening questionnaire consisting of five items mnemonically represented by the acronym SCOFF (Guarda & Redgrave, 2004). Items of the SCOFF refer to eating patterns involving loss of control and feeling sick, weight loss over the last three months, self-perception of being fat, and pervasiveness of food in one's life. Items are scored as one (as opposed to zero) when endorsed and a total score of two and above has been found to detect an eating disorder with very high sensitivity (1.00) and specificity (.88; Morgan et al., 1999).

Eating Disorder Screen for Primary Care (ESP)

The ESP, like the SCOFF, is a brief screener of eating disorders intended for use in primary care settings (Cotton, Ball, & Robinson, 2003). It consists of five questions regarding: the respondent's satisfaction with their eating patterns; whether they eat in secret; whether their weight influences their self-esteem; their family history of eating disorders; and whether they have had or currently have an eating disorder (Guarda & Redgrave, 2004). Using a cut-off score of two abnormal responses, the ESP has demonstrated high sensitivity (1.00) and specificity of (.71) for an eating disorder diagnosis (Cotton et al., 2003). A comparison of the ESP with the SCOFF, using the Questionnaire for Eating Disorder Diagnosis as the independent standard, has found the ESP to have higher sensitivity but lower specificity than the SCOFF (Cotton et al., 2003).

South Oaks Gambling Screen (SOGS)

The SOGS is a 20-item questionnaire based on DSM-III criteria for pathological gambling and was designed to be self-administered or administered by an interviewer (Lesieur & Blume, 1987). The SOGS covers a broad range of areas including type of gambling, symptoms and negative consequences of problem gambling, and sources of money for gambling. The SOGS is the most used instrument in gambling research literature, though use has recently declined due to criticisms of it being overly-sensitive. In a study with a gambling treatment sample and a general population sample, the SOGS demonstrated satisfactory validity by displaying high correlations with DSM-IV diagnostic criteria and moderate correlations with other measures of gambling problem severity (Stinchfield, 2002). The SOGS displayed good to excellent classification accuracy in the gambling treatment sample, but poorer accuracy in the general population sample, with a tendency to overestimate the number of pathological gamblers as compared to DSM-IV criteria (Stinchfield, 2002). For this study, a preliminary screening question ("Have you gambled in the last 3 months?") was added and participants who responded "yes" to this question were administered the SOGS.

4.2.4 Procedure.

Online Questionnaire.

Within seven days following their consent to the study, participants were sent an email inviting them to complete the Online Questionnaire via a hyperlink included in the email. To commence the questionnaire, participants were prompted to enter their e-PASS username. One hundred and seventy eight participants (21.7% of overall included participants) completed the Online Questionnaire, and it is unknown how many commenced but did not finish it. The estimated completion time was approximately 15 minutes based on trials conducted by the author. On several occasions during the data collection period, a reminder email was sent out to recently consented participants who had yet to complete the Online Questionnaire.

Clinical Interview (CI).

Clinical interviews were conducted over the telephone. The use of telephone interviewing for assessing mental disorders has support in the literature (Aziz & Kenford, 2004; Evans et al., 2004; Kobak, Williams, Jeglic, Salvucci, & Sharp, 2008; Rohde, Lewinsohn, & Seeley, 1997). Four interviewers were involved: two provisional psychologists completing a Masters/Doctorate degree in clinical psychology, and two fully qualified psychologists. An interview protocol was

developed to enhance consistency in interview administration. All interviewers received training and supervision for the interviewing process and were experienced in the schedules used. Interviewers attempted to contact participants by phone within four weeks of their e-PASS completion. One hundred and fifty eight participants were reached and agreed to the CI, while four declined due to personal reasons (e.g., too busy or no longer interested). Interviewers were blind to participants' e-PASS results.

At the beginning of the telephone call, background information was provided regarding confidentiality and an overview of the interview process. The CI then proceeded with the MINI-Plus schedule. If a participant expressed any degree of current or past anxiety symptoms, the MINI-Plus was followed with the corresponding anxiety disorder modules of the ADIS-IV. For example, if a participant reported having symptoms of anxiety attacks, as initially queried in the panic disorder module of the MINI-Plus, the interviewer later administered the panic disorder module of the ADIS-IV.

Broad screening questions around difficulties with excessive eating, sleeping, or gambling were then asked and participants who responded positively were presented with the EDE, Insomnia Severity Index, or DIGS, respectively. When deemed by the interviewer to meet criteria for a disorder, participants were asked to discuss, then rate the associated severity of distress and interference on 9-point Likert scale (0 = "not at all", 4 = "moderate", 8 = "very severe"). After administering all of the relevant schedules, interviewers asked additional questions to clarify any diagnostic uncertainties where needed. For a few participants who raised more severe risk issues (e.g., suicidal ideation) interviewers assessed in greater detail, offered the option to cease the interview, providing relevant referral information, and encouraging follow-up consultation with a health professional. At the end of each CI, participants were provided a brief summary of their reported symptoms during the interview but did not receive a formal diagnosis. Participants were then invited to participate in the user experience related studies (see Chapters Five and Six) and thanked for their participation.

Following each CI, interviewers completed an interview rating form (see Appendix E) to summarise interview data for a participant. On the form, interviewers listed diagnoses made according to the MINI-Plus. The participant's severity ratings (on a 9-point Likert scale, 0 = "not at all", 4 = "moderate", 8 = "very severe") for any disorders reported during the interview were also detailed. In addition, interviewers were required to consider all available information (e.g., from all of the schedules and additional questioning) collected from the interview and, applying their clinical judgement, write down their diagnosed disorders with their perceived rating of severity,

using the same 9-point Likert scale presented to participants. If an interviewer decided that there were significant symptoms for a particular disorder, but insufficient to meet diagnostic criteria, the interview labelled it as “subclinical”. The interviewer was required to write down any additional information that pertained to diagnosis and rated their level of confidence in their diagnostic summary. Finally, the interviewer noted the interview completion time.

The rating form also required interviewers to rate: participants’ level of suicidality as either “nil”, “low”, “moderate”, or “high”, and; participants’ degree of agreement with their e-PASS results on a five-point Likert scale (1 = “completely disagree”, 5 = “completely agree”). Finally, specific interview details were also recorded on the form including minimal identifying details of the participant, interview duration, and information to assist in research administration. Two activities were performed to measure the inter-rater reliability of interview results. Firstly, a subset of interviews were recorded and reviewed by a different interviewer who was blind to the initial results. Secondly, interview documentation (e.g., rating form, completed assessment schedules, and any additional notes) for a randomly selected group of participants was reviewed by an experienced clinician, blind to the initial interview results.

Test-retest reliability sub-study: Repeating the e-PASS.

One hundred participants were randomly selected from the total participation pool between June 2010 and June 2011 to be involved in the Reliability Sub-Study. Each of these participants was sent an email within 48 hours of consenting to the overall study, inviting them to participate in the reliability study by repeating the e-PASS. To do so, participants were required to reply to the email briefly stating whether or not they intended to repeat the e-PASS, and then click on a hyperlink provided in the invitation email. This took participants to a special access point in the e-PASS, which bypassed the terms and conditions and initial background questions usually presented to first-time users of the e-PASS. However, participants were required to enter the original e-PASS username to allow identification. A total of 60 participants repeated the e-PASS and all retest data was stored alongside participant’s original e-PASS data in the student investigator’s data file.

4.2.5 Statistical analyses.

Participants’ background.

Participant background variables (e.g., gender, education, employment status, treatment history) were analysed using descriptive statistics. To examine the representativeness of the

various subsamples (e.g. Interview group, Retest group), the background variables were compared between the subsamples and the total sample (*N* = 616). Pearson's chi-square test was used to compare differences between categorical background variables (e.g., gender) while independent samples *t*-tests were used to compare differences in continuous variables (e.g., mean number of diagnoses). Standard deviations were calculated for all mean values and statistical tests were generally deemed significant if resulting in a *p* value of less than .05 (two tailed) or otherwise stated.

Correlation of diagnoses.

The association between e-PASS diagnoses (absent, subclinical, clinical) and diagnoses made by the CI was measured using either Phi or Cramer's *V* (Field, 2009). Phi is used when the association involves two nominal variables, each with two categories. Cramer's V is used for measuring the association between two nominal variables when one of these variables has three or more categories. Cramer's *V* can vary from .00 to 1.00 and the following can be used to interpret the strength of association (Lea & Parker, 1997): .01 to .05 = negligible relationship; .06 to .10 = weak relationship; .11 to .15 = moderate relationship; .15 to .25 = strong relationship; .25 or greater = very strong relationship.

Validity and reliability.

Ruland et al. (2007) recommend that as a minimum standard, research instruments require evidence of at least one type of reliability for the population group being tested, one type of content validity, and at least one type of criterion-related or construct validity. As a form of content validity, the face validity of the e-PASS is evident in the resemblance between e-PASS items and DSM-IV-TR criteria. Face validity was also supported during program development, when clinical psychology staff reviewed and approved of e-PASS items and underlying branching/diagnostic rules. For this study, test-retest reliability was the focus of reliability testing while criterion validity was investigated by analysing concurrent and predictive validity.

Criterion Validity.

Criterion validity is a measure of how well results from one instrument correspond with or predict concurrently administered external measures conceptually related to the measured construct. The diagnostic results of the CI was considered the "gold standard" and therefore used as the criteria for assessing the validity of the e-PASS's diagnostic results. Consistent with the diagnostic thresholds applied in the e-PASS and CI, a severity rating of 3.5 or greater was

deemed a positive clinical diagnosis for a disorder. The CI also recorded subclinical diagnoses. This was given when a participant either: endorsed most but not all symptoms to meet diagnosis; or endorsed all symptoms but did not report sufficient associated distress or impairment as reflected in a severity score of below 3.5 for a particular disorder.

Diagnostic classification statistics.

The diagnostic accuracy of the e-PASS was examined by calculating several diagnostic classification statistics. Conditional probabilities used consisted of sensitivity, specificity, positive/negative predictive values, and positive/negative likelihood ratio values. The *sensitivity* of the e-PASS refers to the proportion of people with an actual disorder (i.e. according to the CI) who are diagnosed by the e-PASS as having the disorder. In contrast, *specificity* refers to the proportion of people without a disorder who are correctly diagnosed by the e-PASS as not having the disorder. Both sensitivity and specificity range from .00 and 1.00, with higher values generally indicating better accuracy.

There are no generally agreed upon levels of acceptable sensitivity and specificity, as the choice may vary depending on the aim, costs and benefits of a diagnostic tool (Smit, Smit, Cuijpers, & De Graaf, 2007). However, in line with other validation studies of online diagnostic assessment programs (e.g., Donker et al., 2008; van Ballegooijen et al., 2012), a sensitivity of .70 or above was considered acceptable in the current study. As a good diagnostic tool should also have at least moderate ability to rule out a diagnosis, acceptable specificity was also set at .70 or above. This value is higher than what other studies have considered satisfactory for screening purposes (e.g., Donker et al., 2008; van Ballegooijen et al., 2012).

The *positive predictive value* (PPV) is defined by the probability of actually having a disorder given a positive diagnosis of the disorder by the e-PASS (Lalkhen & McCluskey, 2008). Alternatively, negative predictive value (NPP) refers to the probability of not actually having a disorder given a negative diagnosis of the disorder by the e-PASS (Lalkhen & McCluskey, 2008). Unlike sensitivity and specificity, PPV and NPP are influenced by the base rate of the disorder being diagnosed, with higher base rates tending to increase PPV and decrease NPV (Garb, Lillienfield, & Fowler, 2008). In general, the PPV and NPV terms are commonly associated with the value of a test for clinicians (Lalkhen & McCluskey, 2008).

The likelihood ratio of the e-PASS is the probability of a particular e-PASS diagnostic result (e.g., clinical diagnosis of panic disorder) in individuals with the disorder (i.e., panic disorder) divided by the probability of the same e-PASS result in individuals without the disorder

(McGee, 2002). In contrast to PPV and NPV statistics, the likelihood ratio is independent of the base rate of the diagnosed disorder in the population (Attia, 2003). A positive likelihood ratio reflects the odds that an individual with a positive e-PASS result actually has a clinical diagnosis of the disorder in question (Attia, 2003). On the other hand, a negative likelihood ratio indicates the odds that a person receiving a negative e-PASS result actually has the clinical diagnosis of the disorder in question (Attia, 2003).

Likelihood ratio values greater than one increase the likelihood of a particular diagnosis. For example, a positive likelihood ratio of two for an e-PASS clinical diagnosis of panic disorder means that individuals who receive this are twice as likely to have an actual clinical diagnosis of panic disorder, compared with someone who does not receive an e-PASS clinical diagnosis of panic disorder. Positive likelihood ratio values above one result in the following approximate increases in probability of having an actual particular clinical disorder: 2 = 15%, 3 = 20%, 4 = 25%, 5 = 30%, 8 = 40%, 10 = 45+% (McGee, 2002); whilst negative likelihood ratio values below one result in the following approximate decreases in probability of having an actual particular clinical disorder: 0.1 = 45%, 0.2 = 30%, 0.3 = 25%, 0.4 = 20%, 0.5 = 15% (McGee, 2002).

The degree of diagnostic agreement between the e-PASS and CIs was measured by Cohen's kappa (Cohen, 1960). This is the most widely used statistic for measuring the proportion of agreement between two categorical variables beyond that expected by chance (Streiner, 2003). Kappa typically ranges from 0 to 1, with zero indicating chance agreement, and unity representing perfect agreement in every case (Sim & Wright, 2005). There are several guidelines for interpreting kappa which vary in the wording of descriptors and thresholds (Sim & Wright, 2005). However, those proposed by Landis and Koch (1977) are most frequently cited and were therefore used in the current study to interpret the strength of agreement between the clinical diagnoses of the e-PASS and CI: .01 to .20 = slight, .21-.40 = fair, .41-.60 = moderate, .61-.80 = substantial, and .81-1.00 = almost perfect agreement.

As a measure of agreement, kappa does not consider whether one variable is a "gold standard" or represents a true diagnosis (Sim & Wright, 2005). When there is disagreement, kappa also does not distinguish whether it is due to random differences (i.e., due to chance) or systematic differences (i.e., those due to consistent pattern; Sim & Wright, 2005). Furthermore, the kappa statistic is influenced by the actual prevalence of the diagnosis considered (Attia, 2003). For low prevalence diagnoses, very low kappa values may not necessarily reflect low rates of overall agreement. Conversely, diagnoses with very high prevalence may inflate the overall

agreement indicated by a high kappa score. Therefore, e-PASS diagnoses with very low/high prevalence and low/high kappa values require close examination.

The relationship between e-PASS diagnostic results and severity was also examined by calculating the bivariate correlation between the number of e-PASS clinical diagnoses with measures of distress and impairment, such as the Kessler-6 scale, and participants' subjective ratings of quality of life. Independent *t*-tests were then calculated to investigate whether a positive e-PASS clinical diagnosis corresponds with higher scores of distress and impairment than a negative diagnosis. Levene's test was used to test the homogeneity of variances assumption in this case. All variables were also screened for normality and outliers prior to analysis to adhere to the assumption of normality when conducting t-tests.

Test-retest reliability.

The agreement between the e-PASS's diagnostic results between test and retest was firstly analysed with Cohen's kappa statistic. This was calculated for each disorder with two or more clinical diagnoses in either the test or retest stage. The permitted time period between test and retest can influence kappa. An interval too brief may allow the respondent to recall and replicate their previous response; and if the interview is too lengthy, the symptoms being examined may change. Given the presumed stability of the disorders targeted by the e-PASS, a 2 to 25 day interval was permitted. The aforementioned Landis and Koch (1977) guidelines for interpreting kappa were also applied here.

McNemar's test was also used to examine any changes in e-PASS clinical diagnosis for a particular disorder between test and retest. With McNemar's test, the diagnostic outcomes of the test are compared against those of the retest in a two-by-two contingency table, and the difference between the marginal portions (e.g., positive diagnosis on initial test, and negative diagnosis on retest) of the table is tested for statistical significance based on a chi-square result (Field, 2009; Sheskin, 2000). A significant result implies the need to reject the null hypothesis that the clinical diagnosis for a particular disorder has remained consistent between test and retest. An examination of the contingency table can then show whether the inconsistency reflects a pattern of change from a positive to negative, or negative to positive diagnosis from test to retest.

Predictive validity using binary logistic regression.

Binary logistic regression was conducted to assess the extent to which e-PASS severity scores (0 - 8) could predict the outcome of a CI diagnosis (1 = present clinical diagnosis, 0 =

absent clinical diagnosis). Binary logistic regression was chosen because of its suitability in modeling relationships with binary outcomes (Field, 2009). Binary logistic regression has one main assumption: that the dichotomous outcome (i.e., a positive or negative CI diagnosis) follows a binomial distribution across the range of predictor values (Peng, Lee, & Ingersoll, 2002), which for this study, are the severity scores (i.e., 0 to 8) of an e-PASS diagnosis for a particular disorder. As the sample for this study was random, in that individual cases and their associated results were independent of one another, this assumption can be considered valid. It is recommended that a sample size for binary logistic regression be at least 100 participants, with a minimum ratio of 10 participants to 1 variable (Tabachnick & Fidell, 2001). The sample for this study met these requirements.

The process of logistic regression modeling involves transforming non-linear data into logarithmic terms ("logit") so that the logit of the outcome variable (presence "1" or absence "0" of a clinical diagnosis) can then have a linear relationship with the predictors (i.e. e-PASS severity score of 0 to 8). The resulting equation for π represents the probability of an outcome occurring and varies between 0 and 1. The predictors in the equation (e.g. e-PASS severity score) each have coefficients (i.e. β) which are estimated by fitting the logistic regression model to the observed data.

Figure 3. The Simple Logistic Regression Model

$$logit(Y) = natural\log(odds) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$$

$$\pi = Probability\ (Y = Outcome\ of\ interest, X = x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$

SPSS version 18.0 was used to perform the logistic regression and default settings were selected ("Enter" method of regression; maximum of 20 iterations). Binary logistic regression is recommended for sample data with at least 20% positive cases (Berry & Linoff, 2000; Garson, n. d.). Therefore, analysis was restricted to the disorders which had at least a 20% prevalence rate of positive clinical diagnosis in the CI: panic disorder (with/without agoraphobia; 22%), GAD (26%); MDD (0.28%); social phobia (36%); and insomnia (23%). The dichotomous outcome variables consisted of either the presence (coded as 1) or absence (coded as 0) of a clinical diagnosis for each of these disorders. A single predictor logistic regression model (LRM) was initially fitted to the data to test whether e-PASS severity scores of a particular disorder could

adequately predict a CI diagnosis for the same disorder. This process was repeated for the five disorders aforementioned.

To address the question of whether a combination of e-PASS severity scores (i.e., of different disorders) could enhance prediction beyond the use of a single predictor (i.e., the corresponding e-PASS severity score), logistic regression modeling was performed with multiple predictors and a step-wise approach. The e-PASS severity scores selected as predictors were those which previously demonstrated a significant correlation with the outcome variable. The Backward Likelihood Ratio (LR) method of logistic regression was selected as it addresses suppressor effects and minimises potential Type II errors (Field, 2009). The Backward LR method initially enters all predictor variables, then tests whether the removal of predictors significantly influences the models fit with the observed data. Predictors were removed one at a time (referred to as "steps") in order of contribution to the LRM, beginning with the predictor of least influence. The amount of contribution by a predictor to model fit was measured by the change in -2 Log Likelihood statistic of the model following the removal of the particular predictor.

Effectiveness of logistic regression model.

The effectiveness of each LRM was assessed using four criteria (Peng et al., 2002): overall model evaluation; goodness-of-fit statistics; tests of the individual predictors; and validation of predicted probabilities. Firstly, the overall model was evaluated by assessing whether the model with a predictor (or several predictors) provided an improvement over the intercept-only model (i.e., the null model). The improvement was examined by three inferential statistical tests: the likelihood ratio, score, and Wald tests. While each test yields similar results, the likelihood ratio was prioritised as recommended when there are differences between these tests (Menard, 1995).

Goodness-of-fit between the predicted outcomes and actual outcomes was measured by the Hosmer-Lemeshow (H-L) test, Cox and Snell's R^2, and Nagelkerke's R^2 (Field, 2009). The Hosmer-Lemeshow (H-L) test is a Pearson chi-square statistic with significant values representing a lack of model fit to the data (Peng et al., 2002). When calculating H-L, the number of groups formed from the estimated probabilities should ideally exceed five; whilst each group should have an equal number of observations and expected frequencies greater than five. The Cox and Snell's R^2 and Nagelkerke's R^2 are variations of R^2 and used as supplementary indicators, with higher values generally indicating better fit between predicted and actual outcomes (Peng et al., 2002).

To test individual predictors, the Wald chi-square statistic and Odds ratio of a LRM's coefficients were examined. The Wald chi-square test indicates whether the intercepts and coefficients of a LRM differ significantly from zero, and therefore, contribute to the prediction of the outcome. The Wald statistic should be used cautiously when a regression coefficient is large, because the standard error may be inflated, thus underestimating the Wald statistic (Menard, 1995). The odds ratio (e^B) indicates the change in odds of a positive diagnosis resulting from a unit increase in the predictor (Field, 2009), which in this case is the e-PASS severity score. As an example, an odds ratio of 4.00 indicates that a one unit increase in the predictor variable (e.g., e-PASS severity score of panic disorder from 3 to 4), corresponds with a four times increase in the odds of a positive diagnosis (e.g., clinical diagnosis of panic disorder). The odds of a positive diagnosis refers to the probability of a positive diagnosis occurring, divided by the probability of a positive diagnosis not occurring (i.e., the absence of a diagnosis, or a "negative" diagnosis).

Finally, the validity of the predicted diagnosis was measured against the observed diagnosis made by the CI (i.e., validity criterion). To calculate a predicted outcome (i.e., presence or absence of a clinical diagnosis), a cut-off point had to be applied to the probability distribution of the predicted LRM. While the default is usually .50, a cut-off point reflecting the prevalence rate of the diagnosis from the CI was instead adopted as recommended by Maloof (2003) to deal with imbalanced data sets. For example, a clinical diagnosis of panic disorder was observed in 22% of the CI sample; therefore, a probability cut-off point of .22 was used to calculate predicted outcomes, with predicted cases falling above the .22 threshold being classified as a positive diagnosis, and those falling below being classified as having an absent diagnosis. Classification statistics (sensitivity, specificity, PPV, NPV, Cohen's kappa) were then calculated for the relationship between the predicted and observed diagnoses.

Receiver Operating Characteristics (ROC) and Area under the Curve (AUC).

Classification statistics of the predicted outcome are influenced by the selected cut-off point. Using a relatively lower probability cut-off point generally increases sensitivity and the positive likelihood ratio, but decreases specificity and the negative likelihood ratio. Therefore, to bypass the subjectivity of selecting a cut-off point, ROC curves were calculated for each of the LRMs. A ROC curve plots sensitivity (or true positive rate) against the false positive rate (or 1-specificity rate) as the cut-off point is varied within the LRM (Lobo, Jimenez-Valverde, & Real, 2008). By calculating the area under the ROC curve (AUC), a single value representing the accuracy of prediction across all possible cut-off points can be generated. In general, the AUC is

considered a standard measure of the overall accuracy of a predictive model (Lobo et al., 2008). Another way of viewing AUC is that it equals the likelihood of a randomly chosen participant with a positive CI outcome scoring a higher predicted value than a randomly chosen participant with a negative CI outcome (Cairney et al., 2007).

Several limitations of the AUC have been highlighted (see Lobo et al., 2008). For example, it does not take into consideration the goodness-of-fit of the prediction model; hence, a poorly fitted model can still have good discrimination and vice-versa (Hosmer & Lemeshow, 2000). Also, the AUC summarises test performance across all cut-off points, but it is unlikely that extremely high or low cut-off points would ever be considered as they correspond with high false-positive and negative rates, respectively. In addition, the AUC places equal weighting to false-positives and false-negatives, when in reality, the two may have different implications (e.g., cost of misclassification). In light of these and other issues, the AUC should be interpreted in conjunction with examining the ROC curve and LRM statistics such as the H-L goodness-of-fit measure. Despite its issues, AUC has been used as a measure of accuracy in a number of studies (e.g., Donker et al., 2008; van Ballegooijen et al., 2012). In this study, AUC values and their confidence intervals of the LRMs were calculated and compared. The following guidelines were used to interpret AUC values: .50-.70 = low accuracy; .70 - .90 = medium accuracy; .90-1.00 = high accuracy (Fischer, Bachmann, & Jaeschke, 2003).

4.3 Results.

4.3.1 Health and treatment background of total sample.

Of the 616 participants (i.e., the total sample) who completed the health and treatment background questions, over 99% reported having a current primary mental health concern (see Table 5 below), with the most common being anxiety. Amongst the 41.1% of the total sample who reported receiving current therapy, general counselling followed by cognitive behavioural therapy were the most frequently undertaken therapy modalities. Approximately 40% of the total sample was using some form of medication at the time of participation, with over half of this group using antidepressant medication.

Table 5

Percentage of Total Sample with Current Mental Health Concerns and Treatment

Current mental health concern	Percentage (%)
None	0.7
Anxiety	59.8
Depression	13.6
Eating / weight issues	13.3
Stress	7.5
Substance/alcohol issues	1.1
Other	4.0
Current therapy	
None	58.9
General counselling	22.2
Cognitive behavioural	14.5
Hypnosis	0.7
Other (e.g., ACT, DBT, mindfulness, psychodynamic)	3.7
Current medication	
None	59.0
Antidepressant (e.g., Zoloft, Effexor)	26.0
Anxiolytic, hypnotics (e.g. Valium, Xanax)	7.7
Antipsychotic (e.g., Sertraline)	3.0
Mood stabiliser (e.g., Lithium)	2.8
Other (e.g., Suboxone, Tegretol)	1.0

Note. $N = 616$

As represented in Table 6, almost half of the total sample was accessing a mental health service (e.g., medical doctor, psychiatrist, psychologist, counsellor) at the time of participation ("current"), whilst over 80% had a history of accessing a service ("ever"). The most popular "current" health profession was a medical doctor, though when considering past or present access, "psychologist" had slightly higher rates of access with approximately 42% of the total sample reporting this. The vast majority of participants had never accessed a mental health

nurse, social, telephone or online service. Psychiatrists and self-help-books were reported with similar rates of use by total participants at some point in the past or present.

Table 6

Percentage (%) of Mental Health Service Access amongst Total Sample

	Current	Past 12 months	Prior to last 12 months	Ever	Never
Access of mental health services	48.2	16.5	17.3	82.0	12.0
Medical doctor	26.0	7.5	7.1	40.6	59.4
Psychiatrist	13.7	2.4	4.1	20.1	79.9
Psychologist	21.5	9.4	11.4	41.9	58.1
Counsellor	5.5	5.2	6.8	17.5	82.5
Mental health nurse, social worker	2.8	1.6	1.8	6.2	93.8
Telephone based service	2.1	1.6	1.6	5.3	94.7
Online service	2.1	2.6	0.8	5.5	94.5
Self-help book	10.1	4.7	6.5	21.3	78.7

Note. N = 616

Participants reported being absent from work due to sickness a mean average of 3.86 days (*SD* = 7.23 days) in the past month. Less than 36% of the sample visited a general practitioner in the past month, with the mean average number of visits being 1.34 (*SD* = 1.55) per month for the total sample. Thirty-nine percent of the total sample reported being diagnosed with a physical health condition, with 11.6% reporting being overweight, 7.8% having diabetes, and 6.5% reporting hypertension.

When asked whether they felt they "have an adequate level of social support or engagement in social/community activities", 58% of the total sample replied in the negative. When asked to rate their overall level of self-confidence in managing their mental health, the majority of participants selected either "poor" (24.7%), "neither poor nor good" (34.2%), or "good" (26.5%). Fewer participants rated "very poor" (9.0%) or "very good" (4.5%). Almost two thirds of the total sample rated their overall quality of life as "neither poor nor good" (30.7%) or "good" (35.5 %), whilst a quarter of the total sample rated either "very poor" (5.6%) or "poor" (18.9%).

Regarding the use of coping strategies when stressed, 28.2% of the total sample reported drinking alcohol, whilst 9.3% reported using substances. Almost a third of the total sample reported exercising (30.7%), whilst almost half of the total sample found talking to friends and family (44.3%) a helpful coping strategy. Only 15.7% of the total sample endorsed "seeing a medical doctor" as a helpful coping strategy when stressed. Approximately 35% of the total sample listed "other" coping strategies which included bingeing and purging, sleeping, self-harm, watching TV, and "writing stuff down".

When asked about their motivation for treatment, 44.3% of participants indicated they were "prepared to take action" whilst a third (33.4%) indicated they were "in the process of making changes". The remainder indicated they were "relapsed and looking for additional assistance" (17.7%), "not interested or needing treatment" (0.5%), or "neither here nor there" (3.0%). Finally, regarding participants' self-identified learning styles, over half indicated they learnt best by doing (51.8%). A quarter of participants responded that they learnt best by reading, while the remainder said they learn best by "looking or watching" (17.0%) or hearing (6.7%).

Diagnostic rates of the e-PASS.

The mean number of total subclinical or clinical diagnoses per participant was 5.38 (*SD* = 2.47, min = 0, max. = 13) out of a possible 21. There were no significant differences in the number of diagnoses between men and women across the total sample. Table 7 (below) shows the e-PASS prevalence rates of subclinical and clinical diagnoses for the total sample. Specific phobia was the most frequently diagnosed disorder, with over 80% of participants receiving either a subclinical or clinical diagnosis. Higher rates of clinical diagnoses were seen for the anxiety disorders, MDD, and insomnia. There were relatively few cases with a clinical diagnosis of a substance disorder, somatisation disorder, and pathological gambling, while the least diagnosed was anorexia.

Table 7
Prevalence Rates (%) of e-PASS Subclinical and Clinical Diagnoses in the Total Sample

	Absent	Subclinical	Clinical
Panic disorder	61.2	13.0	25.8
Social phobia	43.5	26.8	29.7
Specific phobia	15.6	21.6	62.8
GAD	34.1	27.1	38.8
OCD	73.1	19.6	7.3
PTSD	61.2	23.2	12.8
MDD	28.6	30.2	41.2
Anorexia	99.2	0.5	0.3
Bulimia	79.5	9.1	11.4
BDD	80.7	4.1	15.3
BED	93.2	3.4	3.2
Alcohol dependence	79.7	17.9	2.4
Cannabis dependence	92.9	5.7	1.5
Opioid dependence	94.6	4.9	0.5
Sedative dependence	87.5	10.7	1.6
Stimulant dependence	92.9	5.7	1.5
Somatisation disorder	97.9	0.3	1.8
Pathological gambling	94.8	4.5	0.6
Insomnia	32.5	31.0	36.5

Note. N = 616

4.3.2 e-PASS versus the Clinical Interview.

Interview group characteristics.

Of the 616 participants who completed the e-PASS, 158 participants went on to complete the CI (Interview group). Interviews were conducted on average 10.4 days (*SD* = 7 days, min. = 1, max. 34 days) after participants completed the e-PASS, with a mean duration of 48 minutes (*SD* = 15 minutes, min. = 7 minutes, max. = 96 minutes). The Interview group's mean age was 39.74 years (*SD* = 12.0) and did not differ significantly from the mean age of the total sample (*M* = 37.7, *SD* = 12.9). Table 8 presents several sociodemographic characteristics of the Interview group compared against the Total sample. Using a chi-squared goodness-of-fit test, it was shown that the sociodemographic characteristics for the Interview group did not differ significantly from the sociodemographic characteristics of the total sample. However, the Interview group appeared to have slightly higher rates of: metropolitan residents, graduates with year 12 and post-secondary education, and full-time employees.

Table 8

Comparing Demographic Variables of Total Sample and Interview Sub-sample

	Total sample % *N* = 616	Interview group % *n* = 158	Chi-squared* (*df*)
Gender			0.21 (1)
Male	28.1	26.9	
Female	71.9	73.4	
Relationship			0.78 (4)
Married	28.7	28.5	
Single	27.6	28.5	
De-factor	28.3	29.1	
Separated or divorced	10.8	8.9	
Other	5.6	6.3	
Country of birth			3.00 (5)
Australia	73.5	74.1	
United Kingdom	8.7	8.6	
Asian countries	5.1	5.7	
US	3.7	1.3	
European country (except UK)	3.7	3.8	
Other	6.0	6.5	
Setting			2.67 (3)
Metropolitan	62.3	65.9	
Regional	25.2	22.8	
Rural	10.6	8.2	
Remote	1.9	3.2	
Highest schooling			3.74 (3)
Year 9 or less	5.8	4.5	
Year 10	11.4	7.1	
Year 11	6.7	7.7	
Year 12	76.1	80.8	
Highest post-school education			6.03 (5)
None	14.4	10.8	
Current undergraduate	13.4	9.5	
Undergraduate	23.4	24.7	
Postgraduate	19.0	24.1	
Diploma, apprenticeship, trade	14.9	13.9	
Certificate	14.8	16.5	
Employment			2.55 (6)
Full time	38.1	41.0	
Part time	28.4	26.3	
Disability, maternity, sick leave	7.1	6.4	
Home duties/carer	7.0	5.1	
Retired	3.1	4.5	
Unemployed	10.2	10.9	
Other (e.g. volunteer, student)	6.0	5.8	

Note. * All Chi-square tests were insignificant ($p > .05$)

As seen in Table 9 (below), a larger proportion of the Interview group were receiving current mental health assistance compared with the Total sample, and this difference approached significance. Across the individual services, the Interview group generally had only slightly higher rates of access. Similarly, a higher proportion of the Interview group received treatment at the time of participation compared with the total sample. Cognitive-behavioural therapy (CBT) was the only treatment type endorsed by a significantly higher percentage of the Interview group (21%) compared with 14% in the Total sample (χ^2= 6.01, *df* = 1, *p* = .01).

Table 9
Comparing Mental Health Service and Treatment Use between the Interview Group and Total Sample

	Total sample %	Interview group %	Chi-square*
Receiving current mental health assistance	47.7	55.1	3.42
Current service type			
Medical doctor	26.0	28.8	0.67
Psychologist	21.1	26.9	3.17
Psychiatrist	13.6	16.7	1.22
Self-help book	10.1	12.8	1.31
Counsellor	5.5	8.3	2.37
Telephone service	2.1	1.9	0.03
Online service	2.1	2.6	0.16
Social worker	1.8	1.3	N/A
Mental health nurse	1.0	1.3	N/A
Current treatment type			
Antidepressants (e.g. Zoloft, Effexor)	25.8	25.6	0.00
General counselling	23.3	25.0	0.25
CBT	14.3	21.2	6.01
Anxiolytic or hypnotic (e.g. Valium)	8.3	6.4	0.72
Antipsychotic med. (e.g. Seroquel)	3.6	5.1	1.10
Mood stabilizer (e.g. Lithium)	2.8	5.1	N/A
Hypnosis	0.8	1.9	N/A

Note. Total sample *N* = 616, Interview group *n* = 158; * *df* = 1

The mean number of e-PASS diagnosed disorders (either subclinical or clinical) for participants in the Interview group was 5.49 disorders (*SD* = 2.56) and did not significantly differ

from the rest of the sample (*M* = 5.34, *SD* = 2.44, *t* = 0.66, *df* = 614, *p* = .56). The percentages of e-PASS subclinical and clinical diagnoses across the various disorders were generally consistent between the Interview group and the total sample. The rate of subclinical diagnoses of panic disorder in the Interview group was noticeably larger than that seen in the Total sample, however, the difference did not reach significance (χ^2 = 4.89, *df* = 2, *p* = .09). These results therefore suggests the Interview group was a representative sample of the general e-PASS population. The next analysis concerns only the Interview group.

Table 10 depicts basic descriptive statistics of the severity scores for e-PASS and CI diagnoses for this sample of 158 participants. Given that a large proportion of the sample had e-PASS and CI severity scores of zero across the range of disorders, the majority of the severity scores displayed significant skewness and kurtosis ($p < 0.01$). Therefore, a mean comparison (e.g. paired *t*-test) of severity scores between the e-PASS and CI diagnoses was not performed. Instead, non-parametric agreement rankings (kappa) were conducted to examine the association between the main diagnostic categories (i.e. absent/subclinical/clinical) of the e-PASS and the CI.

Table 10
Descriptive Statistics of e-PASS and Clinical Interview Severity Scores

	e-PASS				Clinical interview			
	Min.	Max.	Mean	*SD*	Min.	Max.	Mean	*SD*
Panic disorder	0.00	8.00	1.53	2.25	0.00	8.00	1.22	1.97
OCD	0.00	8.00	0.65	1.51	0.00	8.00	0.51	1.44
GAD	0.00	8.00	2.68	2.40	0.00	6.00	1.58	2.07
Social phobia	0.00	8.00	1.95	2.33	0.00	7.00	1.83	2.06
Specific phobia	0.00	8.00	1.34	2.04	0.00	6.00	0.44	1.17
PTSD	0.00	8.00	1.17	1.91	0.00	7.00	0.49	1.55
MDD	0.00	8.00	2.70	2.35	0.00	7.00	1.38	2.16
Bulimia	0.00	8.00	0.60	1.73	0.00	7.00	0.35	1.34
BED	0.00	4.83	0.25	0.90	0.00	5.00	0.09	0.60
BDD	0.00	8.00	0.70	1.86	0.00	7.00	0.41	1.31
Somatisation disorder	0.00	8.00	0.23	1.22	0.00	5.00	0.11	0.72
Pathological gambling	0.00	2.83	0.14	0.51	0.00	0.00	0.00	0.00
Cannabis dependence	0.00	4.50	0.15	0.55	0.00	6.00	0.07	0.55
Stimulant dependence	0.00	2.33	0.07	0.34	0.00	4.00	0.03	0.33
Opioid dependence	0.00	8.00	0.23	1.04	0.00	5.00	0.07	0.50
Sedative dependence	0.00	6.33	0.23	0.81	0.00	4.50	0.10	0.62
Alcohol dependence	0.00	6.00	0.42	0.98	0.00	7.00	0.30	1.03
Insomnia	0.00	8.00	2.30	2.27	0.00	7.00	1.42	1.86

Note. *n* = 158

Association between e-PASS diagnoses.

The strength of association, as measured by kappa, was measured between the diagnostic categories (absent/subclnical/clinical) of e-PASS diagnoses across the Total sample. The purpose of this analysis was to gauge whether certain e-PASS diagnoses were more likely to co-occur. Associations with significant kappa values ($p < .05$) are presented in Table 11 (below). Anorexia and opioid dependence were not included as they did not have any significant associations. Most of the anxiety disorders as well as MDD and insomnia had significant associations with one another. For instance, panic disorder significantly associated with 12 other e-PASS diagnoses considered, ranging from slight to weak in size (BED, OCD, MDD, insomnia) to fair in size (social phobia, PTSD). The largest kappa value referred to a fair to moderate association between the e-PASS diagnoses of MDD and insomnia (.37), implying that an e-PASS subclinical or clinical diagnosis of one was likely to present alongside the other.

Other relatively significant kappa values referred to fair associations between social phobia and GAD (.30); BDD and bulimia (.29); specific phobia and social phobia (.28); and GAD and MDD (.28). Several diagnoses with low prevalence in the total sample displayed some significant kappa values with other diagnoses; however, the magnitude of the kappa values (e.g., below .05) indicated very low agreement. Amongst the substance related disorders, sedative dependence had a relatively high (though still fair in size) kappa value in association with panic disorder (.10) and specific phobia (.08), as well as with stimulant dependence (.11). The eating disorders did not present any noticeable associations with non-eating disorders. There was a negative relationship between bulimia and BED (-.08), which was expected given they share characteristics but are mutually exclusive in terms of e-PASS diagnosis. Somewhat unexpected was the significant associations between BDD and numerous disorders, including MDD (.13), social phobia (.12), OCD (.10) and insomnia (.10).

Table 11

Kappa Agreement Statistics Between e-PASS Absent/Subclinical/Clinical Diagnoses

	Panic dis.	Social phobia	GAD	Specific phobia	OCD	PTSD	MDD	Ins.	Alc. dep.	Cann. dep.	Sed. dep.	Stim. dep.	Bul.	BED	BDD	Som. dis.	Path. gamb.
Panic disorder	1.00	.22**	.21**	.29**	.11**	.21**	.13**	.18**	-	.02**	.10**	.04*	-	.04	-	.03	-
Social phobia		1.00	.32**	.16**	.08*	.17**	.24**	.24**	-	-	-	-	-	.04*	.12**	.02	-
GAD			1.00	.20**	-	.15**	.28**	.23**	.04	-	-	-	-	.03	.10**	-	-
Specific phobia				1.00	.14**	.15**	.10**	.12**	-	-	.08*	-	-.06	.05	-	.04*	.04
OCD					1.00	.09*	.06*		-	-	-	.05	.07	-	.10**	-	-
PTSD						1.00	.14**	.16**	-	-	-	-	-	.06*	.09*	.03	.04
MDD							1.00	.37**	.05	-	.04*	-	.04	.04**	.13**	.02	.02
Insomnia								1.00	.04	-	.04*	-	-	-	.10**	.03**	-
Alcohol dep.									1.00	.07	-	-	-	-	-	-	-
Cannabis dep.										1.00	.08	-	-	-	-	-	-
Sedative dep.											1.00	.11**	.06	-	-	.05	-
Stimulant dep.												1.00	-	-	.06*	-	.07
Bulimia													1.00	-.08**	.29**	-	-
BED														1.00	.07	-	-
BDD															1.00	.05	-
Somatisation dis.																1.00	-
Path.. gambling																	1.00

Note. N = 616. Only significant kappa values are displayed; values with no "*": p <.05, * p < .01; ** p <.001

Association between e-PASS and Clinical Interview diagnoses.

To investigate the convergent and discriminant validity of e-PASS diagnoses, the association between e-PASS and CI diagnoses was measured across a range of disorders. Table 12 below depicts significant associations ($p < .05$), as measured by kappa, between the diagnostic categories of the e-PASS and the CI within the Interview group. Several disorders (e.g., stimulant dependence, anorexia) had insufficient data to perform kappa analysis and were not represented. Most disorders had multiple significant associations between different disorders of the e-PASS and CI. However, associations were strongest for corresponding disorders, ranging from fair agreement (e.g., specific phobia = .22) to substantial agreement (e.g., cannabis dependence = .68). For example, the e-PASS diagnosis of GAD resulted in a significant kappa with CI panic disorder (.12), social phobia (.21), MDD (.10) and insomnia (0.11), but had the largest kappa value with the CI diagnosis of GAD (0.32). The largest kappa value was between e-PASS and CI diagnosis of cannabis dependence. An inspection of the contingency table between e-PASS and CI cannabis dependence showed the e-PASS correctly identified (according to the CI) 95.3% (143/150) of cases with absent cannabis dependence, 100% (7/7) of cases with subclinical cannabis dependence, and 100% (1/1) of cases with clinical cannabis dependence. In summary, the results of Table 11 and Table 12 indicate varying levels of association amongst e-PASS diagnoses, and between e-PASS and CI diagnoses. Nevertheless, the strongest associations were between corresponding diagnosed disorders of the e-PASS and CI, which provides preliminary support for the convergent and discriminant validity of e-PASS diagnoses.

Table 12

Kappa Agreement Statistics Between Absent/Subclinical/Clinical Diagnoses of the e-PASS and Clinical Interview

	Clinical interview													
e-PASS	Panic disorder	Social. phobia	GAD	Specific phobia	OCD	PTSD	MDD	Insomnia	Alc. dep.	Opioid dep.	Cann. dep.	Sed. dep,	Bul.	BDD
Panic disorder	.47**	-	-	-	-	-	-	-	-	-	-	-	-	-
Socia phobia	-	.39**	.14	-	-	-	-	-	-	-	-	-	-	-
GAD	.12	.21**	.32**	-	-	-	.10	.11	-	-	-	-	-	-
Specific phobia	.12	-	.17*	.22**	-	.11	-	.20*	-	-	-	-	-	-
OCD	-	-	-	-	.38**	-	-	.12	.15	-	-	-	-	-
PTSD	-	-	-	-	.16*	.32**	.20**	.17*	-	-	-	-	-	-
MDD	-	-	-	-	-	-	.35	.17*	-	-	-	.03	-	-
Insomnia	-	.14	.14	-	-	-	.23**	.35**	-	-	-	-	-	.07
Alcohol dep.	-	-	-	-	-	-	-	-	.53**	-	-	-	-	-
Opioid dep.	-	-	.10*	-	-	.12	-	-	-	.36**	-	.31**	-	-
Cannabis dep.	-	-	-	-	-	-	-	-	.18*	.14*	.68**	-	-	-
Sedative dep.	.10	-	-	-	-	-	-	-	-	-	-	.32**	-	-
Bulimia	-	-	-	-	-	-	-	-	-	-	-	-	.36**	.23**
BDD	-	.10	-	-	-	.14	.14	-	.15*	-	-	-	-	.38**

Note. N = 158. Only significant kappa values are displayed; values with no "*": p <.05; * p < .01; ** p <.001

Clinical diagnostic rates of the e-PASS and the Clinical Interview.

Clinical diagnostic rates and mean severity scores of e-PASS and CI clinical diagnoses were compared. The quantity of clinical diagnoses was positively skewed for both the e-PASS and CIs. Therefore a Wilcoxon Signed-Ranks Test (which does not assume normality) was conducted and found that participants in the Interview group received an average of 2.29 (*SD* = 2.30, min = 0, max. = 10) clinical diagnoses made by the e-PASS which was significantly more than the 1.81 (*SD* = 1.70, min = 0, max. = 7) clinical diagnoses according to the CI (Z = -3.04, p = .001). This indicates the e-PASS produces a higher rate of clinical diagnoses than the CI. Clinical diagnostic rates of individual disorders by the e-PASS and the CI were then examined. Social phobia was the most common disorder diagnosed by the CI, followed by GAD and MDD. Compared to the e-PASS, the CI diagnosed fewer clinical cases across most disorders, especially insomnia, MDD, GAD, BDD, specific phobia, and PTSD. On the other hand, the e-PASS appeared to diagnose a lower rate for only social phobia and OCD when compared with the CI. Pathological gambling, anorexia, and the substance disorders were infrequently diagnosed by both forms.

Table 13
Clinical Diagnostic Rates (%) of the e-PASS and Clinical Interview within the Interview Group

	e-PASS %	Clinical interview %
GAD	42.3	27.2
MDD	40.4	27.6
Insomnia	34.0	22.7
Social phobia	28.2	35.6
Panic disorder	22.4	22.7
Specific phobia	16.0	4.3
PTSD	12.8	8.0
BDD	12.2	7.4
Bulimia	6.4	6.1
Anorexia	0.0	0.0
OCD	5.8	8.6
Somatisation disorder	3.8	2.5
Alcohol dependence	1.9	4.3
Sedative dependence	1.9	1.8
Opioid dependence	1.9	0.6
Cannabis dependence	0.6	0.6

Note. n = 158; stimulant dependence and pathological gambling were not diagnosed by either the e-PASS or CI

A Mann-Whitney *U* test was conducted to compare the average e-PASS severity scores between those with and without a clinical diagnosis by the CI. Stimulant dependence, anorexia, and opioid dependence were excluded from analyses due to insufficient CI diagnoses. As seen in Table 14 (below), the results across the disorders considered were in the expected direction and significant (p <.05). Those with a clinical diagnosis for a particular disorder according to the CI were associated with a significantly higher e-PASS severity score than those with a non-clinical diagnosis by the CI.

Table 14

Mean Rank and Mann-Whitney Test Results of e-PASS Severity Scores as a Function of a Clinical Interview Diagnosis within the Interview Group

	Clinical interview diagnosis		M-W *U* (*df* =158)	*Z*	*p*
e-PASS Severity Scores	Non-clinical	Clinical			
	Mean rank	Mean rank			
GAD	66.33	118.34	806.50	-6.31	**
MDD	62.80	124.15	552.50	-7.60	**
Insomnia	66.52	123.47	613.00	-6.72	**
Social phobia	60.95	114.16	915.00	-7.47	**
Panic disorder	66.33	125.80	532.00	-7.64	**
Specific phobia	77.03	109.86	302.00	-2.09	*
PTSD	74.87	135.79	200.50	-5.02	**
BDD	76.39	121.09	351.00	-5.10	**
Bulimia	75.34	141.10	124.00	-7.05	**
OCD	74.88	127.39	337.50	-5.53	**
Somatisation disorder	78.06	95.38	236.50	-2.27	*
Alcohol dependence	76.95	134.50	143.50	-4.59	**
Sedative dependence	78.55	128.50	85.50	-3.40	**

Note. n = 158; * p < .05; ** p < .001

Primary diagnostic rates and accuracy.

A primary diagnosis refers to the diagnosis of highest severity for an individual, which can be either a subclinical or clinical diagnosis. Primary diagnostic rates by the e-PASS and CI are listed in Table 15 below. Amongst the 158 primary diagnoses (i.e., corresponding to the 158 participants), the most commonly diagnosed according to the CI were social phobia (18.6%), followed by panic disorder (17.9%) and MDD (14.7%). On the other hand, the most frequent e-PASS primary diagnoses were of MDD (17.8%), GAD (16.6%), and insomnia (13.3%).

The accuracy of the e-PASS's primary diagnosis was examined by comparing it against the CI's primary diagnosis (i.e., the validity criteria) for a given case. Almost half of the participants (48.7%, 77/158) received a primary diagnosis by the e-PASS which matched the primary diagnosis by the CI. When considering individual disorders, the highest proportion of correct e-PASS primary diagnoses were for insomnia (100%, 6/6). That is, all six participants with a primary diagnosis of insomnia by the CI also had a primary diagnosis of insomnia according to the e-PASS. Anorexia (100%, 1/1) and cannabis dependence (100%, 1/1) also had high sensitivity, though both had very low prevalence rates. The least accurate e-PASS primary diagnosis was for social phobia (27.6%, 8/29), OCD (33.3%, 2/6) and alcohol dep. (0%, 0/2).

The strength of agreement between the primary diagnoses of the CI and the e-PASS was measured by kappa. As seen in Table 15, there was perfect agreement (1.00) for anorexia and cannabis dependence, though again, the prevalence of these two disorders was very low. There was also substantial agreement in the primary diagnosis of BDD (kappa = .65), while agreement for the remaining primary diagnoses generally varied from fair (e.g., PTSD, social phobia, MDD) to moderate (e.g., specific phobia, bulimia) in size. There was little evidence of agreement for alcohol dependence due to the lack of primary diagnoses.

Table 15

Rate (%) and Agreement (Kappa) of Primary Diagnoses by the e-PASS and Clinical Interview

	CI %	e-PASS %	e-PASS accuracy[(a)]	Kappa
Social phobia	18.6	9.6	27.6	.27*
Panic disorder	17.9	9.6	46.4	.37 *
MDD	14.7	17.8	52.2	.35*
GAD	12.2	16.6	57.9	.31*
Specific phobia	6.4	5.1	40.0	.41*
Bulimia	6.4	7.6	70.0	.47*
PTSD	5.1	4.5	37.5	.19*
OCD	3.8	1.3	33.3	.27*
Insomnia	3.8	13.4	100.0	.29*
BDD	2.6	3.2	75.0	.65*
Alcohol dep.	1.3	0.6	-	-.01
Anorexia	0.6	0.6	100.0	1.00*
Cannabis dep.	0.6	0.6	100.0	1.00*

Note. n = 158; (a) Refers to proportion of CI primary diagnoses that were matched by e-PASS primary diagnosis; * $p < .001$

To examine whether a non-primary diagnosis represented an absence of a diagnosis or a secondary diagnosis (i.e., a subclinical or clinical disorder is present but not the primary diagnosis), three-by-three contingency tables were produced which catagorised primary diagnosis, secondary diagnosis, and absence of diagnosis between the CI and e-PASS. This showed that for many of the higher prevalence disorders, the majority of primary diagnoses made by the CI were either categorised as a primary or secondary diagnosis by the e-PASS. For example, out of the eight participants with a CI primary diagnosis of bulimia, five received an e-PASS primary diagnosis of bulimia and the remaining three received an e-PASS secondary diagnosis of bulimia. This indicates that while the e-PASS may not be perfect in detecting a primary diagnosis, it will at least likely identify an actual primary diagnosis as a secondary diagnosis. Overall, given the large number of disorders considered, the e-PASS displayed a reasonable capacity to accurately determine a primary diagnosis when compared with the CI.

Criterion validity of e-PASS clinical diagnosis.

Of main interest was how well an e-PASS clinical diagnosis agreed with a CI clinical diagnosis as the criterion for validity. For selected disorders, a contingency table of corresponding e-PASS and CI diagnoses and associated classification statistics are represented in Table 16 below. Kappa coefficients were found to be significant at the .05 level across all of the disorders listed. Clinical diagnoses of GAD and OCD by the e-PASS had the lowest kappa magnitudes (.39) and reflect "fair" agreement with the CI. The remaining disorders had magnitudes varying between .47 (bulimia) and .62 (panic disorder), which indicate "moderate" to "substantial" agreement. The e-PASS's sensitivity (i.e. ability to diagnose an actual disorder) ranged from .43 (alcohol dependence) to .86 (MDD), with half of the e-PASS diagnosed disorders falling below the acceptable value of .70. In contrast, specificity (i.e. the ability to diagnose an actual absence of a disorder) varied between .69 (GAD) and 1.00 (alcohol dependence), with most values above .90 and considered very good.

Positive predictive values (PPV) mostly varied between .45 (PTSD) and .77 (social phobia), with the exception of alcohol dependence which had a PPV of 1.00. Negative predictive values (NPV) were more consistent and higher for most disorders, with the smallest magnitude being .80 (social phobia), and the remainder equal to or above .90. From the PPV and NPV results, it appears that an e-PASS clinical diagnosis has a low to moderate likelihood of reflecting a positive clinical diagnosis depending on the disorder, whilst a negative e-PASS diagnosis in general is far more likely to be accurate. Positive likelihood ratios (PLR) were all above 1.00, which indicates that an e-PASS diagnosis is associated with an increased probability of an actual clinical diagnosis. PLR varied from 2.51 (GAD) to 14.60 (bulimia). The PLR of e-PASS bulimia suggests that an individual with a bulimia disorder is 14.60 times more likely to receive an e-PASS clinical diagnosis of bulimia, compared to an individual who does not have the disorder. Furthermore, a PLR value of 14.60 also indicated a greater than 45% increase in the probability of actually having bulimia, when receiving an e-PASS clinical diagnosis of bulimia. Negative likelihood ratios (NLR) varied between 0.17 (MDD) and 0.57 (alcohol dependence). The lowest (and most impressive) NLR was 0.17 for MDD and indicated a 30-45% reduction in the probability of actually having MDD, when receiving an e-PASS clinical diagnosis of MDD.

Table 16

Classification Statistics of e-PASS Clinical Diagnoses against Clinical Interview Clinical Diagnoses

e-PASS diagnosis		Clinical interview		χ^2 (*df* = 1)	Kappa	Sensitivity	Specificity	PPV	NPV	PLR	NLR
		Yes (*n*)	No (*n*)								
Panic disorder	Yes	25	11	60.47*	.62*	.71	.91	.69	.92	7.99	0.31
	No	10	112								
GAD	Yes	32	36	27.27*	.39*	.78	.69	.47	.90	2.51	0.32
	No	9	80								
Social phobia	Yes	34	10	44.90*	.52*	.60	.90	.77	.80	6.02	0.45
	No	23	91								
PTSD	Yes	9	11	45.66*	.52*	.75	.92	.45	.98	9.95	0.27
	No	3	135								
OCD	Yes	5	4	26.80*	.39*	.36	.97	.56	.94	12.86	0.66
	No	9	140								
MDD	Yes	38	24	56.80*	.58*	.86	.79	.61	.94	4.10	0.17
	No	6	90								
Insomnia	Yes	28	22	45.58*	.53*	.78	.82	.56	.93	4.31	0.27
	No	8	100								
BDD	Yes	8	9	42.27*	.51*	.67	.94	.47	.97	10.81	0.36
	No	4	137								
Bulimia	Yes	5	5	33.84*	.47*	.50	.97	.50	.97	14.60	0.52
	No	5	143								
Alcohol dependence	Yes	3	0	68.11*	.59*	.43	1.00	1.00	.97	-	0.57
	No	4	151								

Note. n = 158; * p < .05; PPV = positive predictive value, NPV = negative predictive value; PLR = positive likelihood ratio, NPV = negative likelihood ratio.

Criterion validity of e-PASS subclinical/clinical diagnosis.

Classification statistics were then calculated to determine how well a subclinical or clinical e-PASS diagnosis corresponded with a CI clinical diagnosis. A positive e-PASS result referred to either a subclinical or clinical diagnosis, whilst a negative e-PASS result represented an absence of both. A summary of the classification statistics is presented in Table 17. As expected, many positive cases were identified by the e-PASS, and this translated to improved sensitivity in detecting actual clinical diagnoses across the disorders when compared to the sensitivity of the e-PASS clinical diagnosis alone (in Table 16). That is, sensitivity values were above .85 for most disorders except OCD (.79) and BDD (.74). Sensitivity peaked for MDD, where 98% of those with an actual clinical diagnosis were diagnosed with either subclinical or clinical MDD by the e-PASS. In contrast, specificity values generally decreased from those seen when classification was based on an e-PASS clinical diagnosis. Specificity remained well in the acceptable range (above .70) for several disorders; however, it considerably dropped for GAD (.40), MDD (.37), and insomnia (.42), reflecting the high proportion of false positive results made by the e-PASS of these disorders. The kappa values of the e-PASS subclinical/clinical diagnoses remained significant and ranged from .18 (PTSD) to .47 (panic disorder). Most were considered fair (i.e., .20 to .40) in terms of level of association with a CI diagnosis.

Due to the higher rate of false-positives when including a subclinical e-PASS diagnosis as a positive result, the resulting positive predictive values across the disorders were generally less than those seen when classification was based on the e-PASS clinical diagnosis alone. Only panic disorder and social phobia maintained moderate PPVs, with values of .48 and .57 respectively. As a result of the lower threshold for a positive e-PASS diagnostic result (i.e., subclinical rather than clinical diagnosis), the NPVs accordingly increased for all of the disorders, with the majority above .95. This indicates that an individual with the absence of a relevant clinical disorder is very unlikely to receive a positive e-PASS subclinical or clinical diagnosis for that disorder. The PLR values were generally lower than those seen in Table 16 and varied from 1.54 (GAD) to 7.84 (bulimia). These results indicate the odds of having a CI clinical diagnosis of GAD only increased by 1.54 (or approximately 15% in probability) for individuals with a positive e-PASS subclinical or clinical diagnosis for GAD. Due to the relatively lower number of false negative results, the NLR values were generally better than those seen in Table 16 and ranged from 0.42 (BDD) to 0.06 (MDD). Most NLRs were below 0.20, which indicates at least a 30% decrease in the probability of having the clinical disorder targeted by the e-PASS variable.

Table 17

Classification Statistics of the e-PASS Subclinical or Clinical Diagnoses against Clinical Interview Clinical Diagnoses

e-PASS diagnosis		Clinical interview		Pearson χ^2 (*df* = 1)	Kappa	Sensitivity	Specificity	PPV	NPV	PLR	NLR
		Yes (*n*)	No (*n*)								
Panic disorder	Yes	31	34	41.77*	.47*	.89	.72	.48	.96	3.20	0.16
	No	4	89								
GAD	Yes	37	71	14.06*	.20*	.93	.40	.34	.94	1.54	0.19
	No	3	47								
Social phobia	Yes	48	36	36.91*	.45*	.86	.65	.57	.89	2.43	0.22
	No	8	66								
PTSD	Yes	11	55	13.29*	.18*	.92	.62	.17	.99	2.43	0.13
	No	1	91								
OCD	Yes	11	28	23.99*	.33*	.79	.81	.28	.97	4.04	0.27
	No	3	116								
MDD	Yes	42	72	18.71*	.23*	.98	.37	.37	.98	1.56	0.06
	No	1	43								
Insomnia	Yes	35	71	19.17*	.23*	.97	.42	.33	.98	1.67	0.07
	No	1	51								
BDD	Yes	7	19	19.14*	.31*	.64	.87	.27	.97	4.92	0.42
	No	4	128								
Bulimia	Yes	9	17	42.00*	.45*	.90	.89	.35	.99	7.84	0.11
	No	1	131								
Alcohol dependence	Yes	6	25	20.29*	.26*	.86	.83	.19	.99	5.18	0.17
	No	1	126								

Note. n = 158; *p < .001; PPV = positive predictive value, NPV = negative predictive value; PLR = positive likelihood ratio, NPV = negative likelihood ratio.

Kessler-6 (K-6) association with e-PASS clinical diagnosis.

K-6 scores were analysed to investigate the level of general distress associated with e-PASS clinical diagnoses. Of initial interest was whether K-6 scores associated with the number of e-PASS diagnoses (subclinical or clinical) for an individual. An examination of the variable representing the number of diagnoses made by the e-PASS revealed eleven outliers (each with 12 diagnoses) which were subsequently removed, leaving a total of 604 participants. Resulting skewness (0.12, *SE* = 0.10) and kurtosis (-0.56, *SE* = 0.20) statistics were within an acceptable range (-3.29 to 3.29) whilst the histogram of the number of diagnoses resembled a normal distribution. The K-6 data was then tested for normality and resulted in a skewness value of -0.07 (*SE* = 0.10) and kurtosis value of -0.65 (*SE* = 0.20), with both statistics considered acceptable. An inspection of the histogram indicated an approximately normal distribution and the box plot did not present any additional outliers. A bivariate correlation analysis was then performed, revealing a significant negative association between the number of e-PASS diagnoses and the K-6 scores (R^2 = - .58, p < .001). The size of the correlation indicated a strong relationship, with an increase in the number of e-PASS diagnoses corresponding with higher levels of psychological distress (as reflected by lower K-6 scores).

The overall mean average K-6 score was 18.01 (*SD* = 5.17, min. = 6, max. = 30), which is on the upper limit of the significant distress range (i.e., 6 to 18) associated with serious mental illness. The K-6 scores of all 616 participants were considered and the data appeared normally distributed and free of outliers. Table 18 below presents the mean K-6 scores. An e-PASS clinical diagnosis for a particular disorder resulted in a mean average K-6 score well within the clinically significant distress range, with an e-PASS clinical diagnosis of sedative dependence (*M* = 12.80, *SD* = 4.37) and PTSD (*M* = 13.77, *SD* = 4.48) associated with the lowest average. Alternatively, a non-clinical e-PASS diagnosis for a particular disorder was associated with a mean K-6 score on the upper end of the significant distress range or just outside of it. Those without an e-PASS clinical diagnosis of MDD had the highest mean average K-6 scores (i.e., the lowest average level of distress).

A mean comparison of K-6 scores between those with and without an e-PASS clinical diagnosis was then conducted using *t*-tests. Anorexia, pathological gambling, and certain substance disorders were not considered due to insufficient data. Tests of homogeneity of variances between those with and without a clinical diagnosis were not significant for all remaining diagnoses, except for BDD; for this comparison, a Welch *t*-test assuming unequal variance was applied. For all of the considered disorders, except alcohol dependence, having a

clinical diagnosis was associated with significantly higher levels of psychological distress (as indicated by lower mean average scores), compared with those without a clinical diagnosis ($p < .05$). The largest difference in K-6 scores were seen between those with a clinical diagnosis of MDD ($M = 14.37$, $SD = 4.03$) and those without ($M = 20.62$, $SD = 4.23$; $t = 8.97$, $df = 156$, $p < .001$).

In summary, an e-PASS clinical diagnosis is associated with higher levels of general distress compared with a non-clinical diagnosis. On average, an e-PASS clinical diagnosis falls well within the significant distress range of the K-6, which is further evidence of its clinical significance.

Table 18

The Mean Difference in Kessler-6 Scores between Clinical and Non-Clinical e-PASS Diagnoses

	Clinical diagnosis		Non-clinical diagnosis		*t* value	*p*
	n	Mean (*SD*)	*n*	Mean (*SD*)	(*df* = 614)	value
Panic disorder	159	14.54 (4.74)	457	19.23 (4.75)	10.70	*
Social phobia	183	14.91 (4.65)	433	19.35 (4.80)	10.57	*
GAD	239	15.35 (4.47)	377	19.73 (4.85)	11.21	*
PTSD	96	13.77 (4.48)	520	18.81 (4.90)	9.37	*
OCD	45	14.02 (4.67)	521	18.32 (5.08)	5.37	*
Specific phobia	96	14.85 (4.65)	520	18.60 (5.05)	6.76	*
MDD	254	14.37 (4.03)	362	20.62 (4.23)	18.33	*
Insomnia	225	14.83 (4.38)	391	19.88 (4.67)	13.17	*
Bulimia	70	16.61 (5.18)	546	18.20 (5.14)	2.42	***
BDD	94	14.39 (4.15)	522	18.67 (5.06)	8.86[a]	*
Somatisation disorder	11	14.09 (4.93)	605	18.09 (5.15)	2.55	***
Alcohol dependence	15	15.43 (4.70)	601	18.08 (5.17)	1.90	-
Sedative dependence	11	12.80 (4.37)	605	18.10 (5.14)	3.24	**
Cannabis dependence	9	14.89 (5.16)	607	18.06 (5.16)	1.83	-

Note. $N = 616$; * $p < .001$; ** $p < .01$; *** $p < .05$; [a] *t*-value calculated with equal variances not assumed

Binary logistic regression with a single predictor.

Previous results suggest the 3.5 severity cutoff used by the e-PASS to determine a clinical diagnosis may not be ideal for all disorders. For example, an e-PASS clinical diagnosis of Social phobia was associated with exceptional specificity (.90), but relatively poor sensitivity (.60) when compared against the clinical diagnosis of by the CI. Hence, the question emerged as to whether the severity cutoff for an e-PASS clinical diagnosis should be adjusted to enhance classification statistics. Investigating this required an analysis of ROC curves based on estimated binary logistic regression probabilities for a CI diagnosis.

Single predictor binary logistic regression modeling was performed using the corresponding e-PASS severity scores as predictors. Only five disorders met the criteria of having at least 20% prevalence rate in actual clinical diagnoses: (i.e. panic disorder, MDD, GAD, social phobia, and insomnia) and were therefore of focus. The distribution of related e-PASS severity scores were initially screened for outliers. There were no cases with a Mahalanobis value exceeding the chi-square critical value of 13.82 ($df = 2$, $p < .001$); thus no cases for any of the considered diagnoses were excluded.

The results of the binary logistic regression modeling are presented in Table 19 (below). On the left is a 2 x 2 classification table summarising the frequencies of predicted clinical diagnostic outcomes (i.e. yes/no) based on the calculated logistic regression model (LRM), versus actual clinical diagnostic outcome (as determined by the CI). Comparing observed and predicted frequencies, the sensitivity statistics were found to vary between .47 (GAD) to .72 (social phobia) whilst the specificity measures varied from .86 (social phobia) to .94 (panic disorder and insomnia). By combining the sensitivity and specificity measures, an overall percentage of correct predictions (“gain in certainty”) was calculated for each disorder, and was found to range from 74% (GAD) to 83% (panic disorder).

The Hosmer-Lemeshow (H-L) chi-square statistics were not significant for all five disorders, implying that the calculated LRMs were of at least adequate fit to the data. The Wald statistics were significant ($p < .001$) for all of the slope coefficients (i.e. *B* values) and intercepts, indicating significant contribution to the predicted outcome. The e-PASS’s odds ratios were found to be similar between the disorders, varying between 1.86 (GAD) and 2.35 (MDD). The smallest odds ratio was for GAD, indicating that a one unit increase in e-PASS severity rating for GAD is associated with 1.86 times increase in the estimated odds of receiving a GAD diagnosis by the CI. The largest odds ratio was for MDD, with the odds of a positive MDD diagnosis multiplying by

2.36 with a one unit increase in e-PASS severity rating for MDD. The 95% confidence interviews for each of the odds ratios were all above one (with the smallest lower limit being 1.50 for GAD), indicating that the apparent positive relationship between e-PASS severity scores and positive diagnoses seen in the sample is also likely to be seen in the population (Field, 2009).

Finally, AUC values of the LRMs fell between .83 (GAD) and .89 (MDD) which indicates at least moderate overall accuracy. The 95% confidence intervals suggest the AUC of panic disorder and MDD may each be as high as .95, which is considered highly accurate. An AUC of .95 would indicate that in 95% of cases, a randomly selected individual with an actual disorder would have an e-PASS severity rating for that particular disorder that is larger than that of a randomly chosen individual who does not actually have the disorder (Zweig & Campbell, 1993).

Residuals of the logistic regression models were also examined to identify cases in which the models fitted poorly, and cases that may have overly influenced the models. The residual statistics of focus were: Cook's distance; leverage; standardised and studentised residuals; and DFBeta values. These statistics appeared acceptable for all five regression models.

Table 19

Binary Logistic Regression Analysis of Clinical Interview Diagnoses using e-PASS Severity Scores

	Diagnostic frequencies			Total % correct	Prob. cutoff	Model χ^2 (*df*=1)	Cox Snell R^2	Nag. R^2	H-L test (*p*)	Constant	*B*	SE (*B*)	Wald's χ^2 (*B*) (*df* = 1)	Odds ratio (e^B) (95% CI)	AUC (95% CI)
	Observed	Predicted													
		Yes	No												
Panic disorder	Yes	28	7	83.0	.22	56.83*	.30	.46	4.14 (.26)	-2.77	0.68	0.11	39.03*	1.97 (1.59-2.44)	.88 (.81-.95)
	No	18	105												
GAD	Yes	32	8	72.2	.26	47.84*	.26	.39	8.73 (.12)	-3.21	0.62	0.11	32.01*	1.86 (1.50-2.30)	.83 (.76-.90)
	No	36	82												
Social phobia	Yes	44	14	80.4	.36	58.59*	.31	.43	2.79 (.59)	-1.97	0.65	0.11	38.21*	1.92 (1.56-2.37)	0.84 (.77-.91)
	No	17	85												
MDD	Yes	36	7	81.0	.28	70.69*	.36	.52	2.89 (.82)	-3.93	0.85	0.14	37.50*	2.35 (1.79-3.08)	.89 (.83-.95)
	No	23	92												
Insomnia	Yes	30	6	79.1	.23	52.95*	.29	.43	0.34 (.98)	-3.48	0.72	0.13	33.01*	2.06 (1.61-2.64)	.86 (.79-.93)
	No	27	95												

Note. n = 158; * p < .001; AUC = Area Under the Curve; CI = confidence interval.

Binary logistic regression using multiple predictors.

A further area of investigation was whether logistic regression modeling would better fit the data using a multiple predictor model, including non-corresponding e-PASS severity scores. Logistic regression analysis was repeated on the following disorders: panic disorder, GAD, MDD, insomnia, and social phobia. However, all e-PASS severity scores were included as potential covariates in the regression model and a Backward LR method was applied. Data was initially screened for outliers with no cases removed as Mahalanobis values did not exceed the chi-square critical value on any occasion. Multicollinearity was checked for all variables retained in final predicted models. All tolerance values were far above the minimum threshold of .10; hence, multicollinearity was not deemed a problem.

For the multiple predictor regression modeling of the clinical diagnostic outcome of panic disorder, the regression results suggested a model solely consisting of the e-PASS severity score for panic disorder. Amongst the other predictors considered, e-PASS GAD displayed the most potential as an additional predictor. When this was included as an additional predictor to e-PASS panic disorder, the overall model fit slightly improved in terms of a reduction in -2 Log Likelihood value by 1.44; however, this change was not statistically significant (p = .23). Furthermore the inclusion of the GAD regression coefficient resulted in a non-significant Wald value and Odds ratio below 1.00, together indicating that the variable did not meaningfully add to the prediction of a clinical panic disorder diagnosis.

For the logistic regression modeling of the clinical diagnostic outcome of GAD, the final model included e-PASS severity scores for GAD and social phobia as predictors. The associated regression statistics, as listed in Table 20, indicate that this two predictor model provided a slightly better fit to the observed data (-2 Log likelihood = 127.21, $\chi^2(2) = 51.58$), compared to the model with GAD as the sole predictor (-2 Log likelihood = 130.95, $\chi^2(2) = 47.84$). In the two predictor model, the e-PASS GAD variable showed a far greater contribution to the model, with a larger Wald statistic and Odds ratio, compared to the e-PASS social phobia variable. The e-PASS social phobia variable could arguably be removed, as its regression coefficient in the model had a Wald statistic which was just outside of significance (p = .052), whilst the Odds ratio was reasonably small (1.22, 95% confidence interval of 1.00-1.49).

Logistic regression modeling of the clinical diagnostic outcome of MDD suggested a two predictor regression model consisting of e-PASS severity scores for MDD and insomnia as the predictors. Compared to the previous model with one predictor (e-PASS MDD), the two predictor

model produced a better fit, as seen in the lower -2 Log likelihood statistic and higher Model Chi-Squared, Cox and Snell R^2, and Nagelkerke R^2. In the two predictor model, the e-PASS MDD variable was the superior predictor variable, as it resulted in a more statistically significant Wald value and higher Odds ratio, compared with the e-PASS insomnia predictor variable. However, the AUC statistic of the two predictor model only produced a minor improvement in AUC compared with the single predictor model.

For the clinical diagnostic outcome of insomnia, the multiple predictor logistic regression modeling resulted in a model with three predictors: e-PASS insomnia, GAD and PTSD. The three predictor model improved all of the model fit variables compared to the one predictor model. Each of the three predictors had statistically significant ($p < .05$) Wald chi-squared values for their respective coefficients in the regression equation, with e-PASS insomnia severity scores contributing most to the model. The direction of the slope coefficients suggested that higher severity scores on e-PASS insomnia and PTSD increased the likelihood of a CI diagnosis of insomnia, whilst higher severity scores on e-PASS GAD appeared to reduce the likelihood of a CI diagnosis of insomnia.

For the regression modeling of the clinical diagnostic outcome of social phobia, a three predictor model was initially suggested which consisted of the predictors: e-PASS social phobia, panic disorder, and GAD. This model produced a better overall fit compared with the single predictor (i.e. e-PASS social phobia) model, as seen in the decrease in -2 Log likelihood statistic of 6.40 from that of the one predictor model. The three predictor model also appeared to be more reliable in predicting a clinical diagnosis of social phobia as reflected in the higher Model χ^2 value. Out of the three regression variables, e-PASS social phobia produced a slope coefficient with the highest Wald value and odds ratio. Interestingly, the slope coefficient of the panic disorder predictor variable was -0.23 and had an Odds ratio of 0.79. This suggests that a higher e-PASS severity score results in a lower likelihood of a CI social phobia diagnosis. The third predictor, e-PASS GAD, had a small slope coefficient associated with a small Odds ratio (1.22) and Wald statistic just outside of significance ($\chi^2 = 3.73$, $df = 1$, $p = .053$). Removal of the e-PASS GAD variable resulted in a small reduction in the overall fit of the model which was short of significance (change in -2 Log Likelihood = 3.68, $df = 1$, $p = .055$). Thus it is arguable as to whether the model should include the e-PASS GAD variable.

In summary, the inclusion of additional predictors lead to some improvement in the regression model fit to the observed data for the clinical diagnoses of GAD, MDD, insomnia, and social phobia; and no improvement for the clinical diagnosis of panic disorder. However, there

was little change in the overall prediction accuracy as seen in the minor difference in AUC values between the single and multiple predictor models. In all cases, the AUC for the single predictor model was well inside the 95% confidence interview for the AUC of the multiple predictor models. Hence, there is limited evidence to suggest that the multiple predictor models would result in better overall accuracy than the single predictor models.

Table 20

Logistic Regression Analysis of Clinical Interview Diagnoses using Multiple e-PASS Disorder Severity Scores as Predictors

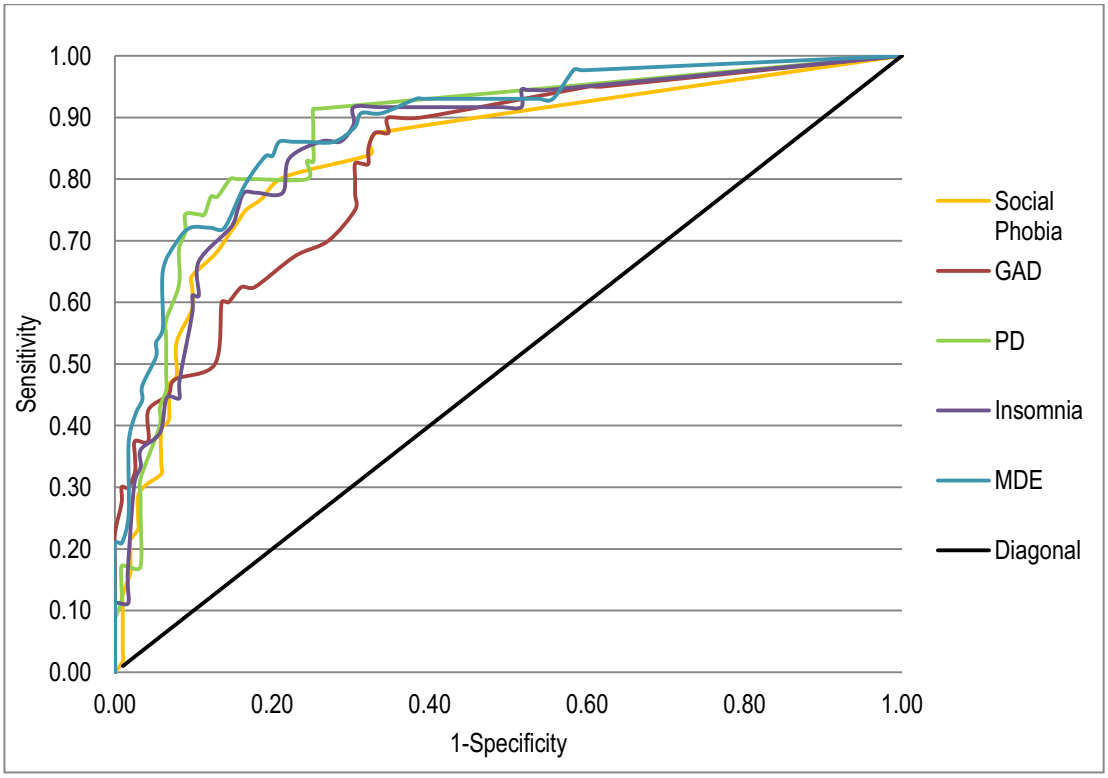
No. of predictors		Regression Variables	*B*	SE (*B*)	Wald's χ^2 (*B*) (*df* = 1)	*p*	Odds ratio (e^B)	95% CI odds ratio	Model $\chi^2(1)$	-2 Log Likelihood	Cox and Snell R^2	Nag. R^2	H-L test (*p*)	AUC (95% CI)
GAD	2	GAD	0.53	0.12	20.58	*	1.69	1.35 - 2.12	51.58*	127.21	.28	.41	3.84	.85 (.78-.91)
		Soc. phob.	0.20	0.10	3.77	.052	1.22	1.00 - 1.49						
		Constant	-3.31	0.49	45.69	*	0.04							
	1	GAD	0.62	0.11	32.01	*	1.86	1.50 - 2.30	47.84*	130.95	.26	.39	8.73	.83 (.76-.90)
MDD	2	MDD	0.72	0.15	24.10	*	2.05	1.54 - 2.73	77.19*	107.79	.39	.56	8.96	.90 (.85-.96)
		Insomnia	0.30	0.12	6.17	.013	1.35	1.07 - 1.71						
		Constant	-4.31	0.65	43.52	*	0.01							
	1	MDD	0.85	0.14	37.50	*	2.35	1.79 - 3.08	70.69*	114.30	.36	.52	2.89	.89 (.83-.95)
Insomnia	3	Insomnia	0.81	0.16	26.89	*	2.25	1.66 - 3.06	62.20*	107.39	.33	.49	12.07 (.15)	.88 (.81-.95)
		PTSD	0.33	0.13	6.80	.009	1.39	1.09 - 1.80						
		GAD	-0.24	0.12	4.07	.043	0.79	0.63 - 0.99						
		Constant	-3.57	0.58	37.80	*	0.03							
	1	Insomnia	0.72	0.13	33.01	*	2.06	1.61 - 2.64	52.95*	116.63	.29	.43	0.34	.86 (.79-.93)
Social phobia	3	Soc. phob.	0.66	0.12	30.13	*	1.93	1.53 - 2.44	63.76*	141.69	.33	.46	10.76 (.22)	.86 (.80-.92)
		Panic dis.	-0.23	0.11	4.29	.038	0.79	0.64 - 0.99						
		GAD	0.20	0.10	3.73	.053	1.22	1.00 – 1.50						
	1	Soc. phob.	0.65	0.11	38.21	*	1.92	1.56 – 2.37	58.59*	148.09	.31	.43	2.79 (.59)	.84 (.77-.91)

Note. n = 158; * p < .001; AUC = Area Under the Curve; CI = confidence interval; H-L = Hosmer-Lemeshow.

Severity cutoff analysis.

Analyses of the Receiver Operating Characteristic (ROC) curves (see Figure 4) derived from the single predictor logistic regression models were conducted to identify e-PASS severity cutoffs which optimised the specificity and sensitivity of the predicted outcomes. This firstly involved examining the coordinates of the ROC curves, which depicts sensitivity against specificity (i.e. 1-specificity) as a function of predicted probability. In general, coordinates to the left of the curve refer to points with relatively low probability thresholds, low sensitivity and high specificity. Coordinates to the right of curve involve higher probability thresholds, and refer to relatively high sensitivity and low specificity. As seen across all five ROC curves in Figure 4, increases in sensitivity beyond .90 were generally associated with considerable larger decreases in specificity, as represented by the gradual plateau towards the right of each curve. On the other hand, increases in specificity above roughly .80 were associated with relatively large decreases in sensitivity, as seen in considerably drop on the left of each curve.

Figure 4. ROC Curve of Binary Logistic Regression Models



Given this trade-off between sensitivity and specificity, it was decided that the optimum probability point on each ROC curve would be chosen based on the maximum sum of sensitivity and specificity, whilst ensuring sensitivity and specificity were at least .70 and .65 in value, respectively. The resulting probabilities associated with the optimal points were then entered into the previously derived single predictor logistic regression equations for each of the targeted disorders (see Table 19. *Logistic Regression Analysis of e-PASS CI Diagnoses*) to calculate the corresponding e-PASS severity value. Table 21 below presents these optimised severity values acting as cut-off scores for e-PASS clinical diagnosis and the classification statistics resulting from them. The current severity cut-off scores and associated classification statistics are also presented for comparison.

Table 21

Comparing Classification Statistics of the e-PASS using Optimised versus Current e-PASS Severity Thresholds

e-PASS diagnosis		e-PASS severity score	Kappa	Sum of sensitivity and specificity	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Positive likelihood ratio	Negitive likelihood ratio
Panic disorder	Optimised	2.25	.59	1.65	.80	.85	.48	.96	3.20	0.16
	Current	3.50	.62	1.62	.71	.91	.61	.91	7.99	0.31
GAD	Optimised	1.75	.42	1.55	.90	.65	.48	.96	3.20	0.16
	Current	3.50	.39	1.47	.78	.69	.47	.90	2.51	0.32
Social phobia	Optimised	1.58	.56	1.58	.80	.78	.48	.96	3.20	0.16
	Current	3.50	.52	1.50	.60	.90	.77	.80	6.02	0.45
MDD	Optimised	3.42	.58	1.65	.86	.81	.48	.96	3.20	0.16
	Current	3.50	.58	1.64	.84	.80	.61	.94	4.10	0.17
Insomnia	Optimised	3.09	.51	1.61	.83	.78	.48	.96	3.20	0.16
	Current	3.50	.53	1.60	.78	.82	.56	.93	4.31	0.27

Note. n = 158

As can be seen in Table 21, the sum of predicted sensitivity and specificity was generally optimised when the e-PASS severity threshold was reduced from 3.5 to between 2.25 (i.e., panic disorder) and 3.09 (i.e., insomnia). The exception was in the case of MDD, where an attempt to identify the optimal threshold resulted approximately to the current threshold of 3.5 and minor changes to classifications statistics. For the other four disorders, the reduction in severity cutoff score resulted in: increases in predicted sensitivity ranging from .09 (panic disorder) to .20 (social phobia) which were generally larger than the resulting decrease in specificity; a small decrease in positive predictive values, except for GAD; an improvement in negative predictive values; a fall in the positive likelihood ratio, except for GAD; and a general improvement in negative likelihood ratio. In summary, there appeared to be small improvements in the e-PASS's clinical diagnosis of panic disorder, social phobia, and insomnia when the severity threshold for determining clinical diagnosis was reduced. The most notable changes were seen in the e-PASS diagnosis of GAD, where a decrease in the severity cut-off from 3.5 to 1.75 lead to improvements in most of the classification statistics considered. In contrast, the results suggest the existing cutoff of 3.5 is optimal for the e-PASS's diagnosis of MDD, and therefore does not warrant change.

4.3.3 Comparing the e-PASS and Online Questionnaire.

One hundred and seventy three participants (Questionnaire group) undertook the Online Questionnaire a mean average of 9.04 days (*SD* = 10.25) after the e-PASS. There were no significant chi-squared differences between the various socio-demographic characteristics of the Questionnaire group and the Total Sample. Hence the Questionnaire group can be viewed as representative of the Total sample. Descriptive statistics of the questionnaire scores are presented in Table 22 below. Only the CES-D and EDE displayed normal distribution whilst the remaining questionnaires generally appeared positively skewed.

Table 22

Descriptive Statistics of the Online Questionnaire

Measure (disorder of focus)	Scoring range	Mean (*SD*)	Median	Min. – max.	Skewness	Kurtosis	Normal distribution
PDSS (Panic disorder)	0-28	8.11 (6.71)	7.00	0 - 28	0.81 (0.19)	0.36 (0.37)	No
GAD7 (GAD)	0-28	10.37 (0.43)	10.00	0 - 21	0.04 (0.19)	-0.99 (0.37)	No
CES-D (MDD)	0-60	28.81 (9.71)	29.00	5 - 52	0.09 (0.19)	-0.41 (0.37)	Yes
SPIN (Social phobia)	0-68	25.43 (15.84)	24.00	0 - 68	0.55 (0.19)	-0.27 (0.38)	No
FQ - Soc. Phob. (Social phobia)	0-40	13.99 (9.20)	13.00	0 - 40	0.58 (0.19)	-0.18 (0.38)	No
FQ –total score (Specific phobia)	0-160	33.97 (25.23)	28.50	0 - 136	1.33 (0.19)	2.01 (0.39)	No
Y-BOCS (OCD)	0-40	8.13 (0.67)	7.00	0 - 31	0.86 (0.19)	-0.37 (0.37)	No
IESR (PTSD)	0-75	15.63 (23.28)	0.00	0 - 85	1.37 (0.19)	0.72 (0.39)	No
ESP (Bulimia)	0-5	2.03 (1.06)	2.00	0 – 4	0.16 (0.19)	-0.66 (0.38)	Yes
SCOFF (Bulimia)	0-5	1.32 (1.48)	1.00	0 - 5	0.89 (0.19)	-0.29 (0.38)	No
BDDQ (BDD)	0-10	3.03 (3.15)	2.00	0 - 9	0.55 (0.19)	-1.16 (0.38)	No
AUDIT (Alcohol dep.)	0-40	7.50 (6.38)	5.00	1- 26	1.10 (0.23)	0.31 (0.45)	No

Note. n= 173

Independent sample *t*-tests were used to compare the mean CES-D scores of group with, and the group without an e-PASS clinical diagnosis (i.e., an e-PASS severity score of 3.5 or greater) of MDD, and to compare the mean ESP score of those with and without an e-PASS clinical diagnosis of bulimia. The assumption of equal variances for the two groups, as tested by Levene's test, was supported for all variables. As depicted in Table 23, participants who received an e-PASS clinical diagnosis of MDD were associated with significantly higher mean scores ($p <$ .001) on the CES-D compared with those who did not receive an e-PASS MDD diagnosis. Furthermore, the group with e-PASS MDD had a mean score above the recommended cutoff score of 27 associated with MDD. Similarly, those with an e-PASS clinical diagnosis of bulimia had a mean ESP score that was significantly higher than the mean for individuals who did not receive a bulimia diagnosis. The mean for the e-PASS bulimia group was higher than the two point cutoff recommended to distinguish an eating disorder, whilst those without an e-PASS bulimia diagnosis had a mean score just below this cutoff.

Table 23

Mean Differences in Questionnaire Scores as a Function of e-PASS Clinical Diagnosis of Target Disorder

Validating questionnaire (Score range)	Target disorder	Quest. cut-off	e-PASS clinical diagnosis				*t* statistic (*df*)
			Yes		No		
			n	*M* (*SD*)	*n*	*M* (*SD*)	
CES-D (0-60)	MDD	27	67	33.00 (9.75)	102	26.07 (6.35)	-4.83* (167)
ESP (0-5)	Bulimia	2	21	3.00 (0.84)	144	1.89 (1.02)	-4.77* (163)

Note. ** $p <$ .001

For the remaining measures with a non-normal scoring distribution, a Mann-Whitney *U* test was employed to measure median rank differences in scores between those with and without a relevant e-PASS clinical diagnosis (Table 24). Significant median rank differences were seen across all of the questionnaires between the e-PASS clinical and non-clinical groups. Effect sizes in group differences varied from moderate (e.g., Y-BOCS, and AUDIT) to large (e.g., PDSS, SPIN, IES). The largest effect sizes between those with and without an e-PASS clinical diagnosis were for the PDSS (i.e., panic disorder) and IES scores (i.e., PTSD). For example, individuals

with an e-PASS clinical diagnosis of panic disorder had a median average PDSS score of 15, whilst those with a non-clinical e-PASS panic disorder diagnosis had a median PDSS score of 6.

For all questionnaires with recommended cut-off scores to differentiate a clinical disorder, the median questionnaire scores for those with and without a corresponding e-PASS clinical diagnosis fell into the appropriate scoring range. For instance, a cut-off score of eight and above has been suggested to distinguish panic disorder with optimal sensitivity and specificity (Shear et al., 2001). In this study, the group with, and the group without an e-PASS diagnosis of panic disorder had a median score of 15 and 6, respectively. This comparison could not be performed for the FQ and BDDQ questionnaires as recommended or standardized cut-off scores were not available.

Overall, an e-PASS clinical diagnosis had at least moderate effect sizes on certain questionnaire scores. The group differences in questionnaire scores seen in the above results further validate the ability of the e-PASS to discriminate significant differences in relevant clinical symptoms. On average, an e-PASS clinical diagnosis is associated with a questionnaire score consistent of a clinical disorder as recommended by previous guidelines.

Table 24

Median Differences in Questionnaire Scores as a Function of e-PASS Clinical Diagnosis of Target Disorder

Validating Questionnaires (score range)	Target disorder	Questionnaire cut-off	Clinical diagnosis on the e-PASS						M-W *U*	*Z*	Effect Size
			Yes			No					
			n	Median	Mean rank	*n*	Median	Mean rank			
PDSS (0-28)	Panic disorder	8.00	35	15.00	140.04	133	6.00	69.88	383.50	-7.63*	.59
SPIN (0-68)	Social phobia	19.00	50	39.00	123.86	114	17.00	64.36	782.00	-7.39*	.57
FQ – Soc. Phob. (0–40)	Social phobia	N/A	50	21.00	121.14	114	10.00	63.66	850.00	-7.25*	.57
FQ – Total (0-160)	Specific phobia	N/A	31	59.00	118.02	125	24.00	68.70	712.50	-5.44*	.44
GAD-7 (0-28)	GAD	10.00	70	14.00	112.69	99	7.00	65.42	1527.00	-6.20*	.48
IES (0-75)	PTSD	23.00 - 33.00	25	57.00	134.64	132	0.00	68.46	259.00	-7.41*	.59
Y-BOCS (0-40)	OCD	16.00	11	22.00	138.45	156	5.00	80.16	259.00	-3.95*	.31
AUDIT 0-40)	Alcohol dep.	20.00	4	22.00	54.57	108	5.00	54.57	7.50	-3.28*	.31
SCOFF (0-5)	Bulimia	2.00	21	4.00	144.57	144	1.00	74.02	219.00	-6.34*	.49
BDDQ (0-10)	BDD	N/A	23	7.00	137.07	143	1.00	74.88	412.50	-5.96*	.47

Note. * Significant at $p < .001$; M-W *U* = Mann-Whitney *U* test statistic.

Logistic regression analysis of the Questionnaires versus the e-PASS.

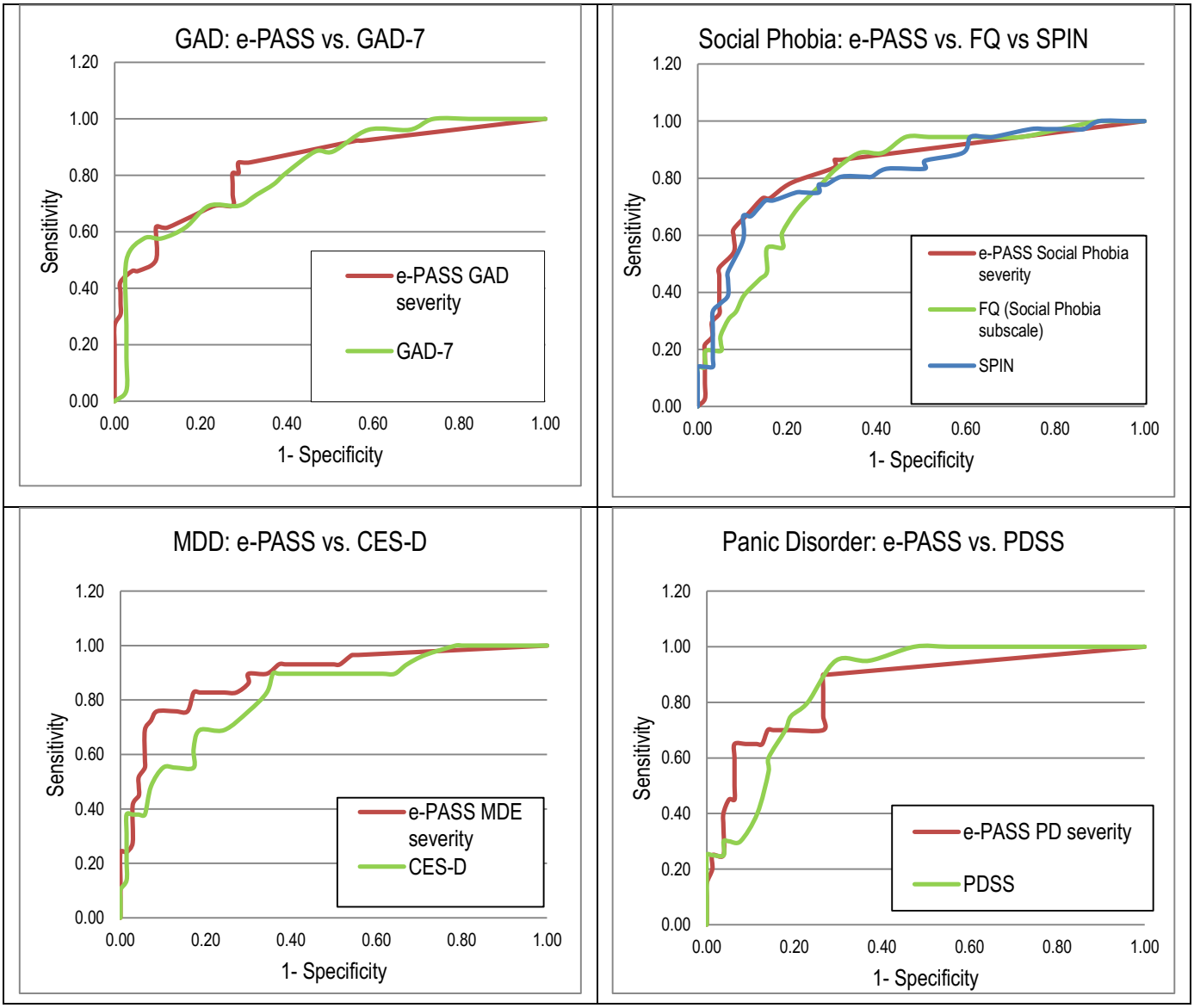
Of further interest was how the questionnaires compared with the e-PASS in terms of predicting a clinical diagnosis by the CI. This investigation involved binary logistic regression analysis and focused on panic disorder, MDD, social phobia, and GAD, given there was comparable data for these disorders across the e-PASS, CI, and Online Questionnaire, and the actual prevalence of these disorders (according to the CI) exceeded 20%. Insomnia was excluded as it was not formally measured by the Online Questionnaire. The sample consisted of 98 participants who completed the e-PASS, the Online Questionnaire, and the CI. The socio-demographic characteristics of this group did not significantly vary from the Total sample according to Chi-square testing. Therefore the group was considered representative of the Total sample.

Using the corresponding questionnaire results as predictors, binary logistic regression analysis using a single-predictor model was performed on the CI diagnostic outcomes of the 98 participants. The process was repeated using e-PASS severity scores as predictors. The logistic regression models involving the e-PASS severity scores had similar characteristics (e.g. Odd's ratios, AUC values) to those produced earlier using a larger sample (n = 158). Statistics regarding model fit and accuracy (AUC) were then compared between models based on the e-PASS severity score and models based on the questionnaires to determine which provided a better fit with the data and was more accurate in predicting CI diagnostic outcomes (Table 25).

For the clinical diagnosis of panic disorder, the regression model based on the PDSS showed similar fit to the data compared with the e-PASS panic disorder model. The e-PASS panic disorder severity score variable was associated with higher Wald statistic and Odd's ratio than the PDSS score variable. However, the overall predictive accuracy as measured by AUC was virtually equivalent. The most notable differences were seen between the CES-D based model and the e-PASS MDD severity based model for predicting a clinical diagnosis of MDD. The model fit variables and coefficient statistics (e.g., Wald and Odd's ratios) of the e-PASS MDD model were much larger than those of the CES-D based model. This suggests that the e-PASS MDD model provides a better fit to the observed data, and that increases in e-PASS severity scores more greatly increase the odds of having a MDD clinical diagnosis, compared with increases in CES-D scores. The overall accuracy of the CES-D model (AUC = .82) was less than that of the e-PASS MDD model (AUC = .89), though there was overlap in the 95% confidence interval. In general, the overall accuracy of the e-PASS based models appeared similar to the

accuracy of the Questionnaire based models. A visual representation of the AUC and the relationship between sensitivity and "1- specificity" for the various regression models are depicted in the ROC curves of Figure 5.

Figure 5. ROC Curve of e-PASS Severity Scores and Questionnaires



An inspection of the curves suggests the e-PASS can achieve moderate (e.g., .60 - .80) sensitivity in diagnosing MDD while maintaining high specificity (e.g., above .90). To reach similar levels of sensitivity, the CES-D appears to sacrifice more specificity than the e-PASS, which is represented by a flatter curve. Similarly, the e-PASS appears more capable than the PDSS in ruling out panic disorder with high levels of specificity and moderate sensitivity. However, when detecting panic disorder with moderate sensitivity, the e-PASS has a greater drop in specificity

than the PDSS. Regarding the diagnosis of social phobia, the e-PASS and SPIN presented a similar potential for sensitivity and specificity, and both result in a greater accuracy than the FQ-social phobia subscale (especially when sensitivity is between .00 and .80), as reflected in the larger AUC values. Finally, the e-PASS seems to maintain a higher level of sensitivity than the GAD-7 when considering specificity levels of .60 to .70.

Enhancing prediction by combining e-PASS severity scores and questionnaire items.

To examine whether the predictive accuracy of the e-PASS severity scores could be improved, step-wise logistic regression was performed using both the e-PASS severity score and individual items of the corresponding questionnaire as potential predictors. For example, the e-PASS GAD severity score and the seven GAD-7 items were inserted as covariates in the logistic regression modelling of GAD clinical diagnostic outcomes (according to the CI). A step-wise approach using a Backward LR method resulted in the removal of six GAD-7 items. The suggested final regression model for GAD included e-PASS GAD severity and item one of GAD-7 ("Over the past 2 weeks, have you been feeling nervous anxious or on edge?") as the predictor variables, and the regression model statistics were generally better than model statistics associated with a model involving the e-PASS or GAD-7 scores alone.

Similarly, the prediction of a clinical diagnosis of MDD using the e-PASS MDD severity scores was enhanced with the inclusion of CES-D items 3 ("I felt that I could not shake off the blues with help from my family or friends"), 7 ("I felt that everything I did was an effort), 12 ("I was happy"), and 14 ("I felt lonely"). The regression model for panic disorder best fitted the data when including e-PASS panic disorder severity score and PDSS questionnaire item one ("How many full panic attacks did you experience in the past month?"). Finally, a combination of the e-PASS social phobia severity score and item 14 of the SPIN questionnaire ("I am afraid of doing things when people are watching") resulted in the optimal regression model for predicting a clinical diagnosis of social phobia. The regression model statistics are presented in Table 25.

Across all enhanced regression models except for panic disorder, the e-PASS severity score contributed more to the predicted outcome (as indicated by larger Wald statistics) compared with the added questionnaire items; in the case of panic disorder, the PDSS item had a larger Wald value than the e-PASS panic disorder severity variable. However, the additional questionnaire items generally had larger Odd's ratios, indicating that an increased score on these items resulted in a larger increase in odds of a clinical diagnosis, compared with the influence on odds resulting from changes in e-PASS severity scores.

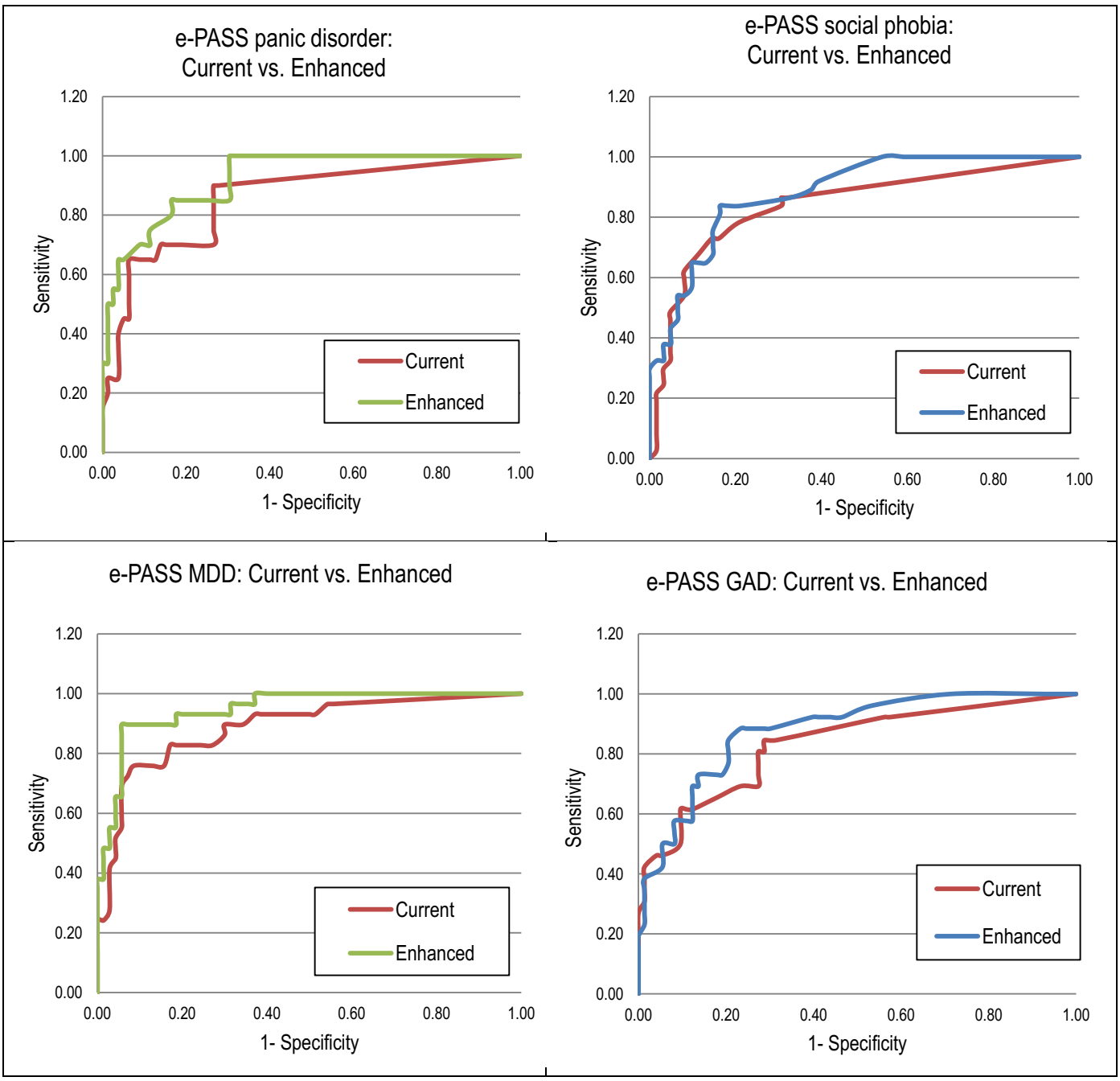
Finally, the enhanced models were associated with considerably larger AUC values than the AUC of the models using solely the e-PASS severity scores or questionnaire totals (see Figure 6). For instance, the AUC of the enhanced panic disorder model was an impressive .92, with a 95% confidence interval (.87 – .98) above the AUC of the regression model involving the e-PASS severity score alone (AUC = .86). Furthermore, compared with the current e-PASS only model, the enhanced models were associated with higher levels of sensitivity at most levels of specificity. These results suggest that the accuracy of the e-PASS could be generally improved by including these additional items in the diagnostic process.

Table 25
Comparison of Logistic Regression Models of Clinical Diagnoses using Questionnaires and e-PASS Severity Scores as Predictors

		Model χ^2 (1)	-2LL	Cox and Snell R^2	Nag. R^2	H-L test (Sign.)	Regression model variables	*B*	SE (*B*)	Wald's χ^2 (*B*) (*df* = 1)	Odds ratio (e^B) (95% CI)	AUC (95% CI)
Panic disorder	PDSS	29.09	70.09	.26	.40	7.37 (.40)		0.24	0.06	18.04*	1.28 (1.14-1.43)	.87 (.79-.94)
	e-PASS	28.65	70.99	.25	.40	12.31 (.01)		0.58	0.13	21.25*	1.79 (1.40-1.29)	.86 (.76-.95)
	e-PASS PDSS (Q1)	44.00	55.64	.36	.57	8.96 (.11)	e-PASS PDSS Q1	0.38 1.37	0.14 0.38	7.56 12.95	1.47 (1.12-1.93) 3.92 (1.86-8.24)	.92 (.87-.98)
GAD	GAD-7	26.15	87.85	.23	.34	9.18 (.33)		0.24	0.06	18.33*	1.27 (1.14-1.41)	.82 (.73-.91)
	e-PASS	32.97	81.04	.28	.41	4.22 (.65)		0.65	0.14	21.85*	1.91 (1.45-2.50)	.83 (.74-.93)
	e-PAS + GAD-7 (Q1)	40.15	73.85	.33	.49	4.20 (.75)	e-PASS GAD7 Q1	0.54 0.88	0.14 0.35	14.51* 6.29***	1.78 (1.30-2.27) 2.41 (1.21-4.80)	.88 (.81-.95)
Social phobia	SPIN	31.03	95.05	.28	.38	3.12 (.87)		0.09	0.02	20.49*	1.09 (1.05-1.13)	.82 (.73-.91)
	FQ	27.60	97.60	.25	.35	4.97 (.76)		0.14	0.03	18.90*	1.15 (1.08-1.23)	.81 (.72-.90)
	e-PASS	37.59	93.29	.32	.43	4.26 (.37)		0.68	0.14	23.23*	1.97 (1.50 – 2.60)	.84 (.76-.93)
	e-PASS + SPIN14	46.17	83.75	.38	.51	6.14 (.29)	e-PASS SPINQ14	0.46 0.71	0.15 0.25	9.80** 8.43**	1.58 (1.19-2.10) 2.04 (1.26-3.30)	.88 (.82-.95)
MDD	CESD	30.90	88.85	.27	.38	7.15 (.52)		0.15	0.03	20.08*	1.16 (1.09-1.24)	.82 (.73-.91)
	e-PASS	45.67*	74.07	.37	.53	4.71 (.58)		0.80	0.16	25.02*	2.20 (1.62-3.03)	.89 (.81-.96)
	e-PASS + CESD (Q3, Q7, Q12, Q14)	50.64	49.01	.51	.73	5.66 (.69)	e-PASS CESDQ3 CESDQ7 CESDQ12 CESD14	0.82 0.92 0.92 -1.49 0.63	0.23 0.36 0.40 0.52 0.31	12.93* 6.51** 5.22** 8.15** 4.18***	2.28 (1.45-3.57) 2.51 (1.24-5.09) 2.50 (1.14-5.48) 0.23 (0.08-0.63) 1.87 (1.03-3.41)	.95 (.91-.99)

Note. * p <.001, ** p < .01, *** p < .05; CI = confidence interval

Figure 6. Comparing ROC Curves of the Current and Enhanced e-PASS Based Models



4.3.4 Test-retest reliability

Of the 60 participants who repeated the e-PASS, 39 did so within 25 days of previously completing the e-PASS and were included in the reliability analysis. The mean average days between testing and retesting was 7.98 (*SD* = 6.63, min. = 1, max. = 25). Participants received an average of 5.05 subclinical or clinical diagnoses on their first e-PASS attempt, and 4.70 subclinical or clinical diagnoses upon retesting, and the difference was not significant ($t = 1.56$, $df = 38$, $p = .13$). Table 26 shows the cross-tabulation of clinical diagnoses made by the e-PASS between (initial) testing and retesting. It also presents the significance level of McNemar's test, the percentage agreement, and the kappa agreement coefficient between clinical diagnoses from initial and retesting. Due to the small size of the sample, the exact binomial probability of the data was used to calculate McNemar's test (Sheskin, 2000). McNemar's test was not significant ($p > .05$) for any of the e-PASS clinical diagnoses, suggesting that the likelihood of change from non-clinical to clinical diagnosis was similar to that of a change from clinical to non-clinical diagnosis. The chi square *p*-value for McNemar's test varied from .22 (BDD) to 1.00 (bulimia). For BDD, there were five participants who changed from a clinical to non-clinical diagnosis of BDD, and only one participant who changed from non-clinical to clinical. Whilst not significant, this pattern suggests that those who initially receive a clinical diagnosis of BDD are less likely to receive this when repeating the e-PASS.

Table 26

Test-Retest Reliability of e-PASS Clinical Diagnoses

	Test	Retest		Agreement	McNemar's	Kappa*
		Clinical	Non-clinical	%	Test – *p* value	
Panic disorder	Clinical	6 (15.4%)	2 (5.1%)	94.9	.500	.83
	Non-clin.	0 (0.0%)	31 (79.5%)			
Social phobia	Clinical	10 (25.6%)	2 (5.1%)	87.1	1.000	.71
	Non-clini	3 (7.7%)	24 (61.5%)			
GAD	Clinical	11(28.2%)	5 (12.8%)	84.6	.219	.67
	Non-clin	1 (2.6%)	22 (56.4%)			
Specific phobia	Clinical	4 (10.3%)	4 (10.3%)	87.2	.375	.54
	Non-clin.	1 (2.6)	30 (76.9%)			
PTSD	Clin.	4 (10.3%)	3 (7.7%)	89.8	.625	.61
	Non-clin.	1 (2.6%)	31 (79.5%)			
MDD	Clinical	11 (28.2%)	5 (12.8%)	79.5	.727	.57
	Non-clin.	3 (7.7%)	20 (51.3%)			
Bulimia	Clinical	10 (28.2%)	1 (2.6%)	97.4	1.000	.87
	Non-clin.	1 (2.6%)	27 (69.2%)			
BDD	Clinical	7 (17.9%)	5 (12.8)	84.6	.219	.60
	Non-clin.	1 (2.6%)	26 (66.7%)			
Insomnia	Clinical	12 (30.8%)	4 (10.3%)	77.0	1.000	.53
	Non-clin.	5 (12.8%)	18 (46.2%)			

Note. $n = 39$, * all kappa values significant at $p < .001$

Table 26 also reveals highly significant kappa agreement coefficients ($p < 0.01$) for all comparisons. Kappa values were excellent for a clinical diagnosis of bulimia (.87), with 10 out of 11 participants initially diagnosed with bulimia also receiving this when retested. Of the two out of 39 cases showing disagreement in bulimia diagnosis, one participant received a BED diagnosis on the first attempt followed by a bulimia diagnosis on the retest; while the other participant received a clinical bulimia diagnosis on the first attempt, and a subclinical bulimia diagnosis on the retest seven days later. The kappa coefficient was also very high for panic disorder (.83), again with only two cases of disagreement: one case changed from a clinical diagnosis of panic

disorder to clinical agoraphobia (without panic), while the other participant changed from clinical panic disorder to subclinical panic disorder diagnosis following a week between test and retest.

The lowest kappa values were for insomnia (.53), specific phobia (.54), and MDD (.57). The data for these comparisons were examined to identify reasons for disagreement. Out of the eight cases with disagreement in clinical diagnosis of MDD between test and retest, five involved a change from subclinical/clinical to clinical/subclinical (respectively), while only three presented a larger change from absent (neither subclinical or clinical) to clinical, or vice versa. Similarly, there were three cases (out of the nine cases with disagreement) with a change from either absent to clinical or clinical to absent diagnoses of insomnia between test and retest. Curiously, two of these cases of major disagreement performed the retest within two days of the initial test, whilst the third was done after six days. For specific phobia, four of the five cases of disagreement involved a change from a clinical to subclinical diagnosis, while the remaining case was of a change from neither a subclinical or clinical diagnosis, to a clinical diagnosis of specific phobia. In general, many of the disagreements between test and retest reflected changes from (to) a clinical to (from) subclinical diagnosis by the e-PASS.

To examine the level of agreement between test and retest of those with a subclinical diagnosis or above (i.e., a clinical diagnosis), the agreement statistics and cross-tabulation tables were repeated, and a summary is presented in Table 27 (below). Again, McNemar's test was not significant for any of the disorders. The kappa statistic and percentage agreement improved for specific phobia, PTSD, and MDD, but worsened for the other disorders.

Table 27

Test-Retest Reliability of e-PASS Subclinical and Clinical Diagnoses

	Test	Retest		Agreement %	McNemar's test - *p* value	Kappa*
		Sub/clin.	Non-clinical			
Panic disorder	Sub/Clin.	10 (25.6%)	2 (5.1%)	91.7	1.000	.76
	Non-clin.	2 (5.1%)	25 (64.1%)			
Social phobia	Sub/Clin.	17 (43.6%)	3 (7.7%)	89.2	1.000	.74
	Non-clin.	2 (7.7%)	17 (43.6%)			
GAD	Sub/Clin.	19 (48.7%)	5 (12.8%)	82.0	.453	.64
	Non-clin.	2 (5.1%)	13 (33.3%)			
Specific phobia	Sub/Clin.	8 (20.5%)	3 (7.7%)	84.6	1.000	.62
	Non-clin.	3 (7.9%)	25 (64.1%)			
PTSD	Sub/Clin.	12 (30.8%)	1 (7.7)	87.2	.375	.73
	Non-clin.	4 (10.3%)	22 (56.4%)			
MDD	Sub/Clin.	22 (56.4%)	4 (10.3%)	84.6	.687	.67
	Non-clin.	2 (5.1%)	11(28.2%)			
Bulimia	Sub/Clin.	11 (28.2%)	2 (5.1%)	87.2	1.000	.72
	Non-clin.	3 (7.7%)	23 (59.0%)			
BDD	Sub/Clin.	7 (17.9%)	5 (12.8%)	82.0	.453	.55
	Non-clin.	2 (5.1%)	25 (64.1%)			
Insomnia	Sub/Clin.	22 (56.4%)	4 (10.3%)	79.5	1.000	.54
	Non-clin.	4 (8.7%)	9 (23.1%)			

Note. n = 39, * all kappa values significant at p <.001

4.3.5 Summary of validity and reliability analyses.

A summary of the psychometric results are presented in Table 28 below. In conclusion, tests of validity reflected mixed results for the e-PASS. The main criterion for validity was the CI and in comparison to this, the e-PASS displayed variable sensitivity for diagnosing a clinical disorder and for several disorders was deemed unacceptable (below .70). On the other hand, specificity was consistently acceptable and mostly high in value except in the case of GAD. For the later disorder as well as OCD, chance corrected agreement as measured by the kappa statistic appeared only fair, while kappa for the remaining disorders ranged from moderate to

substantial in size. Agreement of primary diagnosis between the e-PASS and CI was generally poor to moderate for the disorders considered. Conversely, test-retest reliability for the tested disorders was mostly very good as indicated by high agreement rates and significant kappa values that were moderate to strong in size. However, test-retest reliability results are limited given the small sample size involved.

The e-PASS was also compared against several clinical questionnaires with results generally supporting the validity of the e-PASS. Firstly, a clinical diagnosis by the e-PASS on average correctly associated with clinically meaningful scores on corresponding questionnaires. Furthermore, these questionnaire scores were significantly higher than those associated with a non-clinical diagnosis by the e-PASS. Similarly, an e-PASS clinical diagnosis was associated with K-6 scores indicative of clinical distress and impairment, and significantly greater than scores associated with a non-clinical diagnosis by the e-PASS.

The predictive validity of the e-PASS was also examined with the application of binary logistic regression. This involved testing the extent of which e-PASS severity scores could predict an actual clinical diagnosis according to the CI. As expected, the resulting regression models showed that for particular disorders, higher e-PASS severity scores significantly increased the likelihood of an actual clinical diagnosis, and that this relationship was strongest for MDD and panic disorder. Logistic regression modeling also suggested that the e-PASS was similar to or slightly more accurate than corresponding clinical questionnaires in predicting a CI clinical diagnosis.

With the aim of improving the accuracy of the e-PASS, several potential changes to how the e-PASS produced a clinical diagnosis were investigated: adjusting the e-PASS's severity threshold for a clinical diagnosis; combining the e-PASS severity scores of different disorders; and combining certain items from the clinical questionnaires with the e-PASS severity scores. The results of the regression modeling suggested that lowering the severity score threshold for some e-PASS disorders could improve the overall diagnostic accuracy with a more ideal balance between sensitivity and specificity. It was also suggested that several items from the validating questionnaires could enhance the predictive validity for particular disorders of the e-PASS.

Table 28
Summary of Main Psychometric Results

	e-PASS disorder module										
	Panic disorder	Social phobia	GAD	PTSD	OCD	Specific phobia	MDD	Bulimia nervosa	BDD	Insomnia	Alcohol dependence
Test-retest reliability Kappa*	Strong	Moderate	Moderate	Moderate	N/A	Moderate	Moderate	Strong	Moderate	Moderate	N/A
Criterion Validity (CI)											
Sensitivity	Acceptable	Unacceptable	Acceptable	Acceptable	Unacceptable	-	Acceptable	Unacceptable	Unacceptable	Acceptable	Unacceptable
Specificity	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable	-	Acceptable	Acceptable	Acceptable	Acceptable	Acceptable
Kappa**	Substantial	Moderate	Fair	Moderate	Fair	-	Moderate	Moderate	Moderate	Moderate	Moderate
AUC	Medium	Medium	Medium	-	-	-	Medium	-	-	Medium	-
Criterion validity (Questionnaires)											
Significant effect size?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Medians falling within appropriate range?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	-	Yes
Criterion validity (K-6)											
Significant effect size?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
Medians falling within appropriate range?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Primary diagnosis Agreement between e-PASS and CI	Fair	Fair	Fair	Slight	Fair	Fair	Fair	Moderate	Substantial	Fair	N/A

Note. * Interpret with caution due to limited sample size (*n* = 39); * * Kappa statistic based on agreement of clinical/non-clinical diagnostic results between e-PASS and CI.

4.4 Discussion.

This study aimed to evaluate important psychometric properties of the e-PASS, namely the criterion validity and test-retest reliability of an e-PASS clinical diagnosis. To the knowledge of the author, it is the first study to examine both the validity and reliability of a web-based clinical assessment program with a sample of actual users performing the program in real-world conditions. In addition, this study is novel for investigating a web-based program which targets a large number of DSM-IV disorders in depth, and provides diagnostic feedback that reflects severity and primary diagnosis. These qualities largely differ from other programs easily accessible on the web as well as those reported in the literature which have tended to represent specific symptom scales or very brief screening instruments. This study extends on preliminary psychometric data gained through pilot testing and previously reported community agreement rates amongst AO users (Klein et al., 2011; 2012). Prior to discussing the psychometric results, the profile of e-PASS users will first be considered to provide context of who are directly influenced by the e-PASS findings in the present and subsequent studies.

4.4.1 e-PASS users.

In general, overall demographics were consistent with observed users of treatment programs alongside the e-PASS at AO (Klein et al., 2011). A large proportion of the e-PASS sample identified themselves as having anxiety concerns, being female, and in a relationship. These results are partially consistent with epidemiological findings showing that anxiety disorders are significantly correlated with being female and of middle age (McEvoy et al., 2011). There were also similarities between e-PASS users and respondents of another web-based clinical screening program (MACSCREEN) for common mental disorders (Van Ameringen et al., 2010) in terms of disorder prevalence, average age (35-39 years), proportions reporting anxiety concerns (60%) and previous treatment experience (approximately 50%). Both programs are openly available on web-based “anxiety clinics” and located in similar geographic regions which suggests a typical profile for these circumstances. However, a popular web-based screening program for depression (ISP-D; Lin et al., 2007) in Taiwan has received users that are, on average, 10 years younger and far more likely to be single (82%) than e-PASS users. These demographic differences highlight how online clinical assessment programs may vary in appeal between websites, locations, and population groups.

4.4.2 Criterion validity (Clinical Interview).

Overall, the e-PASS results reflected previous findings indicating that a web-based clinical program can achieve at least adequate criterion validity for the diagnosis of mental disorders (Donker et al., 2009; Farvolden et al., 2002; Lin et al., 2009). However, the e-PASS's diagnostic classification indicators (kappa, sensitivity, specificity, PPV and NPV) varied between disorders, a result also observed in other evaluations of web-based programs (Donker et al., 2009; Farvolden et al., 2002). Out of the eleven e-PASS disorders considered, panic disorder and MDD had the strongest criterion validity and reliability results, and in general, displayed consistently acceptable psychometric properties. Three other e-PASS modules (GAD, insomnia, and PTSD) produced reasonably acceptable results, while the remaining disorders (e.g., social phobia, OCD, bulimia, BDD, alcohol dependence) were associated with one or more unacceptable psychometric properties.

In general, the e-PASS was found to have fair to substantial diagnostic agreement (as measured by kappa) with the CI. Agreement levels resembled those of the WB-DAT (Farvolden et al., 2002) and were far better than diagnostic agreement achieved by the web-based CIDI-SF (Carlbring et al., 2002) against a structured interview. The ability of the e-PASS to detect a clinical disorder (sensitivity) was considered unacceptable for some disorders (e.g., OCD and social phobia), which was somewhat surprising given the high levels of sensitivity that have been previously reported (Donker et al., 2009). On the other hand, e-PASS specificity was generally high, indicating that the e-PASS can identify most non-clinical cases. This quality has been similarly noted in some programs (Farvolden et al., 2002) but not others (Carlbring et al., 2002; Donker et al., 2009).

The precision of an e-PASS diagnosis was also measured with predictive values. Results suggested that a positive e-PASS result is correct in at least 45% of cases, while a negative e-PASS result is almost certain to be correct across most disorders. Other programs have also shown to produce a negative diagnostic results with high precision (Donker et al., 2009; Farvolden et al., 2002), suggesting that web-based programs can be particularly effective at ruling out a clinical disorder. Likelihood values confirmed that a clinical or non-clinical e-PASS diagnosis was associated with an increased probability of that actual result. The e-PASS severity scores were also found to be a good predictor of clinical diagnosis, and that the best predictor for a clinical diagnosis was the corresponding e-PASS severity score. In summary, the e-PASS can be a reasonable good indicator of an actual disorder as diagnosed by a structured CI.

To further appreciate the relative performance of the e-PASS against similar web-based programs, Table 29 presents the classification statistics of clinical disorders that have been mutually evaluated by the e-PASS, WB-DAT (Farvolden et al., 2002) and WSQ (Donker et al., 2009) programs. These programs were also compared against structured interviews and have a similar breadth of disorders.

Table 29

Comparison of Predictive Validity Statistics between e-PASS, WB-DAT, WSQ[a], and CIDI-SF

		Kappa	Sens.	Spec.	PPV	NPV	AUC
MDD	e-PASS	.58	.86	.79	.61	.94	.89 (.83-.95)
	WB-DAT	.68	.79	.89	.75	.93	-
	WSQ[b]	-	.85	.59	.51	.89	.84 (.77-.90)
	CIDI-SF	.00	.25	.76	-	-	-
Panic disorder	e-PASS	.62	.71	.91	.69	.92	.88 (.81-.95)
	WB-DAT[c]	.57	.75	.94	.52	.98	-
	WSQ	-	.90	.44	.38	.97	.76 (.59-.93)
	CIDI-Long[d]	.48	.76	.74	-	-	-
Social phobia	e-PASS	.52	.71	.91	.69	.92	.84 (.77-.91)
	WB-DAT	.59	.74	.94	.56	.96	-
	WSQ	-	.72	.73	.40	.91	.72 (.62-.82)
	CIDI-SF	.24	.73	.62	-	-	-
OCD	e-PASS	.39	.36	.97	.56	.94	-
	WB-DAT	.66	.71	.97	.67	.98	-
	WSQ	-	.80	.69	.15	.98	.81 (.65-.97)
	CIDI-SF	.17	1.00	.85	-	-	-
GAD	e-PASS	.39	.78	.69	.69	.92	.83 (.76-.90)
	WB-DAT	.58	.63	.94	.67	.93	-
	WSQ	-	.93	.45	.29	.97	.77 (.68-.85)
	CIDI-SF	.01	.50	.53	-	-	-
PTSD	e-PASS	.52	.75	.92	.45	.98	-
	WB-DAT	.70	.95	.93	.60	.99	-
	WSQ	-	.83	.47	.11	.99	.65 (0.51-0.80)
	CIDI-SF	-	-	-	-	-	-

Note. (a) WSQ statistics based on refined WSQ (Donker et al., 2009); (b) WSQ measure of any depressive disorder; (c) WB-DAT measure of panic disorder with/without agoraphobia; (d) web-based panic disorder module from standard CIDI

Of the known programs reported in the literature, the WB-DAT most closely resembles the e-PASS in terms of presentation, item structure, and diagnostic scope. As expected, the two demonstrate comparable diagnostic classification statistics, except for certain disorders where the WB-DAT appeared superior in sensitivity (e.g., OCD, PTSD), specificity (i.e., GAD) and overall

diagnostic agreement. However, there were notable methodological differences between the present study and that of the WB-DAT which could explain differences in performance. Firstly, the WB-DAT sample was recruited from clinic research trials and presented lower levels of psychopathology and co-morbidity compared to sample in the present study. For instance, the criterion interview identified panic disorder in 22% of the e-PASS sample as opposed to 9% in the WB-DAT study. With relatively simple cases and less diagnostic issues associated with more severe presentations (e.g. differential diagnosis, clarifying co-morbidity, and categorising symptoms near diagnostic thresholds), diagnostic classification may have been easier, therefore facilitating more diagnostic agreement in the WB-DAT study than in the present study. Also, the WB-DAT study used the SCID as the criterion, which differs from the criterion used in this study, and is perhaps more closely aligned with the structure of the WB-DAT. Future research could directly compare the relative performance of the e-PASS and WB-DAT with the same sample and validity criteria.

On the other hand, the WSQ and CIDI-SF target a similar range of disorders as the e-PASS but have relatively few items given their aim of being very brief screening programs (Carlbring et al., 2002; Donker et al., 2008). Compared with the WSQ and CIDI-SF, the e-PASS displayed similar or lower sensitivity and NPV. Otherwise, the e-PASS produced consistently higher specificity, PPV, and overall accuracy according to AUC values than the WSQ. These results are not surprising given that brief screening instruments such as the WSQ prioritise sensitivity (at the expense of specificity) to maximise detection of potential cases for follow-up assessment (Donker et al., 2008). However, the e-PASS produced far less false-positive results than the WSQ, as reflected in the better PPV. This is likely because the e-PASS has a cumulative item structure requiring endorsement of multiple criteria before assigning a clinical diagnosis, whereas the WSQ uses only one item per disorder which is far less restrictive.

By including more diagnostic criteria and producing relatively less false-positives, the e-PASS appears more conservative in assigning a positive diagnosis than very brief web screeners such as the WSQ. The e-PASS also demonstrates greater potential sensitivity and specificity as indicated by the AUC values. However, the e-PASS is also more burdensome for respondent as it is more complicated and time consuming to complete. While the average e-PASS completion time was 13 minutes, the WSQ takes approximately 2 minutes to complete (Donker et al., 2008). On the other hand, the WSQ's lack of overall accuracy suggests it should be followed up with a formal assessment which in itself adds burden on the consumer. Hence, there is likely a tradeoff

between accuracy and burden of follow-up consultation for web-based program, and it would be of interest to compare this via the e-PASS and a less intensive program such as the WSQ.

4.4.3 Criterion validity (Online Questionnaires).

The validity of an e-PASS diagnosis was also examined against standardised questionnaires of symptom severity, as well as the K-6 measure of distress and impairment. As expected, the results showed significant differences in questionnaire scores between those with and without a relevant e-PASS clinical diagnosis. Effect sizes were generally moderate to large, reflecting a considerable difference in questionnaire scores. Furthermore, the group averages of those with and without an e-PASS diagnosis appropriately fell above and below (respectively) the recommended clinical thresholds associated with the questionnaires. These results are consistent with those found by Donker et al. (2009) in showing that a web-based program can distinguish clinically significant symptoms and distress/impairment along notable instruments.

Interestingly, those without particular e-PASS diagnoses still had average questionnaire scores approaching the recommended clinical threshold. This is perhaps because some questionnaires identify a range of symptoms (e.g., CES-D includes symptoms relevant to anxiety and depression), some of which also relate to other disorders prevalent in the sample. Analyses which control for the influence of co-morbid disorders on questionnaire scores could have further explored this. However, there was evidence to support the accuracy of the questionnaires. Using the CI as criterion, the overall diagnostic accuracy according to regression analysis (i.e., AUC) was calculated for certain questionnaires (CES-D, PDSS, SPIN, FQ, and GAD-7) and resembled previously reported AUC values of the CES-D (Donker et al.,2010), SPIN (Connor et al., 2000) and GAD-7 (Donker et al., 2011). Importantly, the e-PASS showed similar overall accuracy to the corresponding questionnaires for the prediction of panic disorder, GAD, and social phobia. The e-PASS appeared to have better overall accuracy for the prediction of MDD than the CES-D; however, this difference was not statistically tested. These results suggest the e-PASS could be used in place of these questionnaires for the purposes of identifying a clinical disorder.

4.4.4 Number of diagnoses and co-morbidity.

Despite lacking in sensitivity, the e-PASS generated more diagnoses and co-morbidity per person than the CI. This difference has also been noted for other web-based programs (Carlbring et al. 2002; Farvolden et al., 2002). Similarly, equivalence studies have found web-based clinical scales resulting in higher mean scores than the original paper-pencil versions of

the questionnaires (e.g., Andersson et al., 2003; Austin et al., 2006). On the one hand, the larger e-PASS diagnostic rates could be partly due to people being more open and honest about their symptoms on the internet. However, it is difficult to clarify this with the results of the present study. Mean severity scores were generally higher for the e-PASS than the CI, but differences were not tested for significance. While clinical diagnostic rates made by the e-PASS were higher than the CI for some disorders, other disorders had comparable rates. Conversely, the CI produced higher diagnostic rates for social phobia and OCD which is surprising given these two disorders are associated with considerable stigma (Fennell & Liberato, 2006; Lysaker, Yanos, & Outcalt, & Roe, 2010; Stengler-Wenzke, Beck, Holzinger, & Angermeyer. 2004) and would presumably be easier to disclose online as opposed to an interviewer. Furthermore, given the voluntary and informed nature of participation, it seems unlikely that participants would significantly vary the information they disclosed between the e-PASS and CI.

However, the lower rate of social phobia on the e-PASS than the CI is somewhat consistent with Joinson's (1999) findings of lower social anxiety ratings on a web-based questionnaire as opposed to a paper-pencil version, and higher levels of social anxiety when non-anonymous. Joinson explained this by suggesting that responding online reduced social desirability which in turn reduced social anxiety, while responding in less anonymous and non-online conditions (e.g., in a phone interview) raised social anxiety. Therefore, in the present study, the CI may have systematically raised social anxiety levels which were reflected in social phobia symptom ratings.

But with regards to co-morbidity, the e-PASS's generally higher rates were perhaps artificially inflated by the lack of exclusionary criteria built into the e-PASS scoring rules. According to the DSM-IV, exclusionary criteria function as a means of establishing diagnostic boundaries (APA, 1994) and in practice, refer to the exclusion of alternative disorders (i.e. differential diagnoses) before diagnosing a particular disorder. The need for exclusionary criteria has been supported by results showing disorders with exclusionary criteria (e.g. the anxiety disorders) are more likely to co-occur than disorders not associated with exclusionary criteria (Boyd et al., 1984; Slade & Andrews, 2002). In this study, this was evident in highly significant kappa associations between certain e-PASS disorders that share exclusionary criteria (e.g. panic disorder, social phobia, GAD). Logistic regression also suggested combinations of e-PASS diagnoses could better predict a particular CI diagnosis, though the improvement appeared small.

The e-PASS was designed with items reflecting exclusionary criteria to deal with differential diagnosis, especially for the anxiety disorders. However, these item responses were

not factored into the diagnostic outcome of the e-PASS and were not available for analysis in the present study. The CI on the other hand required interviewers to consider differential diagnosis as instructed in the MINI-Plus and ADIS-IV interview schedules. Indeed, interviewers reported several instances where they used clinical judgment (e.g. considering the course and function of symptoms) to apply exclusionary criteria, and ruled out what they believed were redundant diagnoses. Such differences to the e-PASS could help explain the relatively high rates of diagnosis and co-morbidity especially of anxiety disorders by the e-PASS compared with the CI. Further investigation could address whether the e-PASS can effectively apply exclusionary criteria and whether this significantly impacts on diagnostic rates and agreement with a CI.

4.4.5 Primary diagnosis.

The e-PASS distinguishes a primary diagnosis when there is co-morbidity. To the knowledge of the author, this is the first study to examine the classification statistics of a primary diagnosis formed by a web-based program. Contrary to expectations, the e-PASS performed poorly in identifying a primary diagnosis consistent with the CI. Exceptions were the e-PASS primary diagnoses of bulimia and BDD. These two showed moderate to substantial agreement with the CI, perhaps because they diagnostically stood out due to their relatively unique symptom profile, noticeably high level of severity, and low co-morbidity within the sample. In contrast, there was low agreement in primary diagnoses for remaining disorders which was likely related to the lack of criterion validity for some disorders, and the high e-PASS and actual co-morbidity making it more difficult to distinguish the primary diagnosis.

Nevertheless, the results imply that the e-PASS severity ratings are not sufficiently accurate to differentiate primary diagnosis in a consistent manner with the CI. Caution should therefore be taken when considering e-PASS primary diagnostic information as it may not be accurate, particularly when there is co-morbidity in e-PASS results. Fortunately, the primary diagnosis made by the e-PASS currently does not have a significant bearing on treatment access and recommendations within AO. While the e-PASS makes a recommendation to prioritise treatment for a primary diagnosis, users are also encouraged to address any secondary diagnoses that are present.

4.4.6 Disorder-specific performance.

While criterion validity was not perfect across all disorders, the e-PASS was particularly poor at identifying OCD, and overall, showed only a fair level of diagnostic agreement with the CI.

This seems to contrast with previous results indicating better diagnostic agreement and/or sensitivity for web-based screening of OCD (Carlbring et al., 2002; Donker et al., 2009; Farvolden et al., 2002), though the base rate of OCD in one study (Carlbring et al., 2002) was too low to produce generalisable findings. On the other hand, Sheehan and colleagues (1997) also observed poor sensitivity for OCD by a self-administered paper-pencil questionnaire (i.e., the MINI-PR), while other studies have also found self-report measures of OCD to be inferior to clinician administered measures (Stewart, Ceranoqlu, O'Hanley, & Geller, 2005).

4.4.7 Improving e-PASS accuracy.

Lowering diagnostic threshold.

In light of the varied e-PASS classification statistics, several investigations were conducted into how e-PASS accuracy could be improved. As also seen in other studies (Di Nardo et al., 1993), diagnostic disagreement in the present study was not so much due to differences in the presence of symptoms, but rather, the measured severity of associated distress and impairment. Specifically, logistic regression revealed that for certain e-PASS disorders, lowering the e-PASS severity score threshold for a clinical diagnosis could improve diagnostic sensitivity while maintaining an adequate level of specificity. The e-PASS diagnosis of social phobia, in particular, was predicted to produce a more acceptable balance of sensitivity and specificity if the threshold was dropped from 3.5 to 1.58.

Classification statistics for disorders that did not undergo logistic regression also suggested that many e-PASS subclinical diagnoses (i.e., severity score below 3.5) represented clinical diagnoses made by the CI. This was especially evident for e-PASS OCD, which displayed considerably better sensitivity when comparing an e-PASS subclinical or clinical OCD result against a clinical diagnosis of OCD by the CI. Indeed, most clinical presentations recognised by the CI (especially for OCD, social phobia, and bulimia) appeared to receive at least a subclinical result by the e-PASS. In the context of AO, where the e-PASS also functions as a referral program, this is an important finding as it means those who receive a subclinical diagnosis are still recommended treatment (e.g. a self-help program when there is one available for the corresponding disorder).

Applying a lower e-PASS severity threshold for social phobia and OCD has support in the literature. In particular, OCD often presents with poor insight (Kishore, Samar, Reddy, Chandrasekhar, & Thennarasu, 2004), which could manifest as lower e-PASS ratings of distress and interference compared with ratings made by a clinician. A recent review of clinical trials and

longitudinal studies noted that OCD symptoms are not always associated with marked distress (Leckman et al., 2010). Furthermore, Andrews et al. (2010) examined epidemiological data and found clinical cases of OCD and social phobia to be associated with relatively low distress and disability (measure by the K-10) compared with other mental disorders. Andrews et al. suggested altering the diagnostic criteria for these two disorders to reflect a lower pre-requisite for distress and impairment. In the present study, an e-PASS diagnosis of OCD or social phobia produced similar average K-6 scores to other disorders, which does not seem to support the findings of Andrews and colleagues. However, the influence of co-morbid symptoms was not controlled for and may have confounded K-6 results in the present study. Nevertheless, there is still evidence to suggest that the criterion validity of an e-PASS clinical diagnosis could be improved by reducing the severity threshold for diagnosis, particularly in the case of social phobia and OCD.

The question remains as to why the diagnostic threshold of the e-PASS was more strict (i.e., higher) than the CI's. One explanation is that the e-PASS and CI have different underlying processes in determining diagnostic thresholds. The e-PASS is programmed to generate a clinical diagnosis when respondent ratings of severity items are on average 3.5 or above out of 8. The developers chose 3.5 as a general threshold across all disorders because its scale anchor represented mild severity of distress and impairment. The CI, however, incorporated different approaches to determining diagnostic threshold as instructed in the MINI-Plus and ADIS-IV interview schedules used for the CI. The MINI-Plus relies on a yes/no item about experiencing significant distress and impairment. The ADIS-IV asks participants to rate their level of symptom severity, distress, and impairment, and then requires interviewer clinical judgment to determine whether or not the overall presentation is of a clinical disorder. Hence, variations arose between the diagnostic process of the e-PASS and CI which were not factored into analyses, but likely influenced diagnostic agreement. Accordingly, differences between e-PASS and CI thresholds for OCD and social phobia may be due to interviewers emphasizing overall symptom profile rather than reported distress and interference as programmed by the e-PASS. In doing so, an individual may have still received an OCD or social phobia diagnosis by the CI even if they gave relatively low ratings of distress and interference to the e-PASS for these disorders.

These differences in diagnostic thresholds raise the issue of whether one approach is better than another. Applying diagnostic thresholds relate to how one interprets the clinical significance criterion in the DSM-IV, or in other words, distinguishes symptoms as causing a significant level of distress and impairment. As symptoms are considered dimensional, determining the cutoff at which symptoms are "clinically significant" to warrant diagnosis and

treatment is contentious (Andrews et al., 2010). Further, different diagnostic thresholds can greatly influence diagnostic prevalence rates (Andrews et al., 2010; Stein, Walker, & Forde, 1994) and can be a source of diagnostic unreliability (Brown, Di Nardo, Lehman, & Campbell, 2001). Hence, it is recommended that diagnostic thresholds be judged by experts (Andrews et al., 2010) who can carefully consider the nature, course and context of symptoms (Spitzer & Wakefield, 1999). It is doubtful as to whether the e-PASS and similar web-based programs can replicate this complex process, and this could explain the less than perfect agreement with structured clinical interviews observed in the literature. On the other hand, the e-PASS as a computerised program offers a degree of reliability which is perhaps lacking in the clinical judgement underlying the application of diagnostic thresholds. Unfortunately, inter-rater reliability of the CI results was not closely examined to allow this comparison.

Combining e-PASS and questionnaire items.

Another approach considered for improving e-PASS accuracy was to combine items of the e-PASS with items of the criterion questionnaires. A similar approach has been undertaken in previous research (e.g., Donker et al., 2008). Here, logistic regression analysis indicated each of the e-PASS modules considered could be enhanced with the addition of certain questionnaire items, and that the resulting outcome would be more accurate than either the e-PASS or questionnaires alone. For instance, the overall accuracy for an e-PASS clinical diagnosis of Panic panic disorder could be enhanced with the added item of "how many full panic attacks did you experience in the past month?" This is a reasonable recommendation given that the e-PASS in its current form does not specifically address recent panic attack frequency, but instead asks if the respondent has had one or more panic attacks in the past month. Also, measuring panic frequency reflects items in gold standard interviews such as the ADIS-IV, and is consistent with literature highlighting panic attack frequency as an essential feature of panic disorder (Abbar, 1996) and a major risk factor for an anxiety disorder (Goodwin et al., 2004).

However, suggested items did not always appear to add much more than what was already covered in the e-PASS items. For instance, the suggested GAD-7 item overlapped with the content of the opening item from the e-PASS GAD module, except that it referred to a shorter time period when asking respondents to endorse whether they recently experienced feelings of nervousness and anxiety. Other indicated items (e.g., the SPIN item: "I am afraid of doing things when people are watching") were more simply worded than e-PASS items of the same content, suggesting that e-PASS items could be phrased in more basic terms to achieve greater accuracy.

It should be noted that while adding certain questionnaire items might improve prediction, it may not enhance accuracy significantly beyond the current e-PASS form. This was indicated by the overlap in the AUC confidence intervals of the enhanced e-PASS and the current e-PASS. Furthermore, there is the issue of how exactly items should be combined with e-PASS severity scores to enhance prediction; for example, the suggested items could replace similar existing e-PASS items or be presented as additional items. As these suggested changes were derived from regression modelling based on existing data, any potential benefits to e-PASS accuracy would need to be confirmed with new data.

4.4.8 Test-retest reliability.

Another valuable psychometric criterion for a diagnostic instrument is adequate test-retest reliability. This study found support for the test-retest reliability of the e-PASS in terms of moderate to high agreement between initial and retest e-PASS clinical diagnostic results. Test-retest reliability was strongest for panic disorder and bulimia; and weakest for social phobia, insomnia and MDD. Importantly, those who did not have diagnostic agreement between test and retest generally reflected small changes in e-PASS severity, as changes from a clinical to the absence of a clinical or subclinical presentation (and vice versa) were rare. In general, these are positive results given that retesting occurred on average eight days, and a maximum of 25 days after initial testing. The results are also consistent with the widely accepted notion and consistent empirical evidence that computerised administration can offer a high level of reliability (Angle et al., 1978; Butcher, 2004; Kobak et al., 1993; Newman et al., 1997).

Only one known study (Lin et al., 2007) has previously reported the test-retest reliability of a web-based clinical measure. In that study, the test-retest reliability of a MDD diagnosis varied between fair (when retesting occurred between 14 to 28 days) to excellent (when retesting occurred within 14 days). In the present study, the e-PASS produced a higher level of reliability for MDD than the lower range of the ISP-D, though e-PASS reliability values were not calculated for different durations between test-retest to allow direct comparison.

For particular disorders, the e-PASS displayed better test-retest reliability than well-regarded structured interview schedules. For example, Lecrubier et al (1997) examined test-retest reliability of the clinician administered MINI with a sample of 43 participants. The MINI's test-retest reliability measure for panic disorder was smaller than that of the e-PASS despite the MINI being retested within far less days (within two days) than the e-PASS (mean average of

eight days). On the other hand, the e-PASS appeared less reliable than the MINI for GAD and especially MDD.

In general, test-retest reliability of the e-PASS was comparable to reported properties of a computer-assisted CIDI (M-CIDI; Wittchen et al., 1998), though with discrepancies across specific disorders. For both programs, the most reliable diagnoses were of panic disorder. GAD and BN were diagnosed with moderate to excellent reliability by the e-PASS, but only fair reliability by the M-CIDI (Wittchen et al., 1998). Otherwise, e-PASS reliability was similar (e.g., for social phobia) or somewhat lower (e.g., PTSD) than the M-CIDI. However, comparisons between the e-PASS and M-CIDI are limited by methodological differences. The M-CIDI was completed by a much younger sample (14 to 28 years), and in a controlled setting (e.g., in a clinic with an interviewer) which could have promoted more consistent responses. Conversely, retesting of the M-CIDI occurred on average 38.5 days after initial testing, which was a far longer duration than in the present study.

Symptom duration can greatly influence test-retest reliability as noted in other reliability studies (Lin et al., 2007; Wittchen et al., 1998). This issue could help explain the relatively low test-retest reliability measures for the e-PASS diagnoses of insomnia and MDD. Compared to other disorders, e-PASS items for these two disorders particularly emphasised the time period of current symptoms so that only those with recent or current symptoms could receive a diagnosis. For example, the insomnia module required symptoms to occur in the past month in order warrant an insomnia diagnosis. However, given that participants could redo the e-PASS up to 25 days after their initial administration, changes in symptom presentation during this extended period could have more easily resulted in a change of diagnosis for insomnia and MDD. Other disorders assessed by the e-PASS were less restrictive on the course and significance of current symptoms. This issue could be explored in future by conducting further test-retest reliability analyses with different time periods or following modifications to the time period of symptoms reflected in e-PASS items.

On the other hand, test-retest reliability of the e-PASS was reasonably high despite potential factors that could have impaired reliability. Firstly, due to the use of branching and skipping rules, the e-PASS produces a diagnostic outcome based on specific sequences of responses; hence, minor response changes between e-PASS sittings can lead to significant diagnostic changes. Also, due to the unproctored nature of e-PASS administration, there were many extraneous factors (e.g., environmental distraction, boredom, intoxication) that could have influenced performance and, therefore, the results of the e-PASS from one sitting to the next.

4.4.9 Limitations.

While this study found support for the criterion validity and test-retest reliability of the e-PASS, there are several limitations to consider. As some diagnostic classification statistics (e.g., sensitivity/specificity) are influenced by base rates, the results of lower prevalence disorders (e.g., alcohol dependence, bulimia) should be interpreted with caution. Test-retest reliability also involved small base rates as well as sample size (i.e., *n* = 39). While participant numbers were similar to other studies (e.g., Lecrubier et al., 1997; Wittchen et al. 1998), further reliability testing with a larger sample is recommended. Finally, the suggested e-PASS changes derived from regression modeling (e.g., reducing the severity threshold or including new items to improve e-PASS accuracy) need to be verified with a new set of data to ensure they do indeed result in psychometric improvements.

Focusing on criterion validity revealed how well the e-PASS performed against commonly used diagnostic instruments. Whether these criteria are in fact valid was not of primary concern in this study. However, there are several criteria-related issues which may have compromised e-PASS validity results. Firstly, several measures in the Online Questionnaire have not been previously validated for online use (e.g., SCOFF) and therefore, their recommended clinical thresholds may have not been appropriate. Furthermore, although the K-6 is considered a good measure of distress and impairment, it is not disorder specific and is more so used as a general measure of serious mental illness (Kessler et al., 2010).

Despite steps taken to maximise the validity of the CI (e.g., the use of valid structured interview schedules, interviewer training, peer/staff supervision), most interviews were conducted by the one interviewer who may have biased CI results, for example, with their interview style, personality, and clinical skills (Pinninti et al., 2003). A subset of interviews were recorded to ensure quality, however, inter-rater reliability was not thoroughly evaluated. Given highly trained interviewers may still differ in interview style (Lewis, 1994) the use of more interviewers in this study may have averaged out any interviewer effects on diagnostic outcome. There is also the question of whether the telephone medium allows equivalence to face-to-face interview. In this study, they were assumed equivalent. However, face-to-face interviewing involves non-verbal cues and has been suggested to yield longer responses to open questions (Bowling, 2005). Conversely, telephone interviewing is likely more convenient and less confronting for participants, and may have elicited more sensitive information than a face-to-face interview (Bowling, 2005).

While the MINI-Plus is popular as a criterion (e.g., Bunevicius et al., 2007; Cuijpers et al., 2008), it's reported psychometric properties are not optimal (Sheehan et al., 1997). Although

highly regarded, the ADIS-IV also has sources of unreliability related to interviewer disagreement about the number, severity and duration of important symptoms related to diagnosis (Brown et al., 2001). A more robust criterion could have utilised the LEAD standard proposed by Spitzer (1983) and referring to experts forming consensus diagnosis based on all available data. However, this too is associated with diagnostic shortcomings (Kranzler et al., 1995), is very intensive, and may not have significantly changed the results of the present study. As widely recognised, psychiatric diagnosis lacks a clear gold-standard measure in part due to issues associated with the underlying classification system (e.g., overlapping disorders, lack of arbitrary diagnostic criteria with little empirical support) and the reliance on clinician judgment and self-report data, both of which are fallible (Meyer et al., 2001).

Although the e-PASS displayed a good level of test-retest reliability, participants could have responded similarly during the retest to maintain self-consistency. Memory and practice effects may have also inflated test-retest reliability, though this issue was likely minimal in the present study given the reasonable duration between test and retest. More importantly, order effects could have influenced reliability as well as validity but were not investigated due to the fixed order of participation activities. Equivalence studies between paper-pencil and online administration have shown significant, though small order effects for several scales (e.g. BDI-II, MADRS-S; Carlbring et al., 2007; Hollandare et al., 2010). Peters and colleagues (1998) also observed more diagnostic discrepancies when a computerised CIDI was performed before rather than after an interviewer administered CIDI. It was suggested that an initial CIDI interview helped contextualise items, facilitating more consistent responses when the computerised CIDI was subsequently administered (Peters et al., 1998).

In the present research, order effects may have been particularly relevant due to participants undertaking the e-PASS and receiving its diagnostic feedback first. This design was intended to allow a more naturalistic context of e-PASS use. However, it is unclear whether receiving e-PASS results first may have influenced participants' perceived symptoms, and consequently, their subsequent responses to the CI, questionnaires, and repeat administration of the e-PASS. For example, as retesting was voluntary, it's possible those who repeated the e-PASS were more accepting of the e-PASS original results than those who did not repeat the e-PASS, and therefore were willing to repeat their pattern of response to questions. Alternatively, participants who disagreed with initial e-PASS results may have been motivated to respond differently in follow-up tasks. Therefore, further evaluation of the e-PASS could remove the e-PASS feedback, at least for a sub-sample, to see if this feedback does indeed influence follow-up

assessment or retest results. In addition, other order effects could be addressed by counter-balancing the order of completing the e-PASS, the CI, and the self-report questionnaires.

4.4.10. Further research.

There were important population groups and aspects of the e-PASS that were not addressed in the current study but could be the focus of further research. Firstly, certain e-PASS diagnoses (e.g., anorexia, somatisation disorder, substance use disorders, pathological gambling) were overlooked due to low base rates in the sample. Hence further psychometric evaluation could prioritise these disorders by recruiting relevant samples. Also, it is unclear how the e-PASS might perform with less or more clinically severe population groups (e.g., primary care patients, the general public), as varying proportions of people clustering around the diagnostic threshold could influence the diagnostic accuracy of the e-PASS.

Psychometric results were based on a sample considered at least partly representative of the current e-PASS user population. However, the e-PASS will likely attract varying consumer groups in future (e.g., due to wider promotion, or changes to the website drawing different target groups) that may differ in psychopathology and, more generally, how they perceive and respond to the e-PASS. Therefore, further psychometric evaluation should employ larger and more diverse samples to gather relevant e-PASS psychometric data. Future users could then be matched with the appropriate reference groups for more appropriate psychometric guidelines (Buchanan, 2003).

Also, further research is needed into the efficacy of web-based measures of diagnostic severity. Disorder specific severity scores are a novel characteristic of the e-PASS and not commonly seen in web-based clinical assessment programs. However, similar approaches are featured in traditional instruments such as the ADIS-IV and the CIDI interviews (Kessler & Ustun, 2004). Disorder specific distress and impairment questions are recommended as a means of comparing severity between diagnosed disorders and prioritising healthcare resource allocation (van der Feltz-Cornelis, Knispel, & Elfeddali, 2008; Kessler & Ustun, 2004). In the e-PASS, the diagnostic severity measures serve several functions, some of which received mixed support in this study and need further investigation. However, it remains unclear whether e-PASS severity is effective in its other key roles of differentiating diagnostic sub-categories (e.g., subclinical, mild, moderate, severe) and informing appropriate treatment. To investigate this, larger samples with more diverse clinical presentations are needed.

Interview observations suggested highly prevalent but low levels of risk in the e-PASS sample. While the e-PASS briefly screens for suicidal ideation, psychometric properties of this function were not evaluated in this research. Recent research suggests that approximately 50% of people with common mental disorders seeking help on the internet experience suicidal ideation (Hemelrijk et al., 2012). Therefore, it would be worthwhile investigating the e-PASS's ability to identify suicidal ideation, given the likely relevance to e-PASS consumers, and the potential impact suicidal ideation may have on follow-up behaviour.

As seen in other web-based programs, the item set can vary in length and content yet still achieve comparable diagnostic results. While the e-PASS was designed to represent all relevant diagnostic criteria it its items, it is unclear whether all items are in fact psychometrically relevant for diagnosis. Further research could investigate the predictive validity of individual e-PASS items in order to retain critical items and remove/modify redundant items from the current item set (Robins & Lee, 1994). In addition, internal reliability and construct validity of items within a disorder module could be examined by analysing responses to the entire e-PASS item set (i.e., without branching rules applied).

4.5. Conclusion.

As the e-PASS is new, is highly accessed, and has limited psychometric evidence, this study aimed to investigate its psychometric potential and indicate areas of improvement for future program iterations. Unlike many previous studies of web-based programs, the e-PASS was examined with a sample of prospective users who performed the e-PASS in a naturalistic setting. In general, the e-PASS was found to produce clinical diagnostic results that had reasonable criterion validity and good test-retest reliability over a brief period of time. These findings are consistent with previous studies supporting the use of web-based programs for the online screening of mental disorders (e.g., Farvolden et al., 2002, Donker et al., 2008; Lin et al., 2009).

However, e-PASS accuracy was found to be limited in terms of diagnostic differences with the CI. Given the two relied on the same diagnostic criteria, the issue as to why diagnostic differences arose was partially addressed. This study found that for certain disorders, the e-PASS diagnostic severity threshold was too high, and therefore needed lowering (especially for social phobia and OCD) to achieve better diagnostic agreement. Logistic regression analysis also found that certain e-PASS disorder modules could result in greater diagnostic agreement with the CI if they incorporated certain items from other questionnaires. On the other hand, the literature has noted a range of attributes associated with online assessment that distinguish it from traditional

assessment. Therefore, other factors could have also influenced response differences, and consequently, diagnostic differences between the e-PASS and the CI, as well as the e-PASS over repeated administrations. For example, participants may have been relatively disengaged from the e-PASS due to personal and environmental factors such as cognitive issues, severe symptoms of psychopathology, environmental distracters, lower motivation or attitudes to online assessment. Furthermore, participants may have also had greater difficulty understanding and responding to e-PASS items in an appropriate manner.

The CI, in contrast, seemed to require a relatively high level of engagement from participants, and allowed participants to thoroughly discuss and clarify items and responses with interviewers. As a result, the CI took considerably longer and inevitably involved more data collection than the e-PASS. On the other hand, interview discussions may have compromised validity by prompting clinical judgment by interviewers, as well as encouraging socially desirable responses from participants. The extent of which these factors impacted on e-PASS and CI performance is uncertain, but are likely of relevance to the diagnostic agreement between the two. Therefore, the following two studies attempted to investigate the influence of some of these factors as part of an examination into the user experience of the e-PASS.

5.0 Chapter Five: The User Experience of the e-PASS (Study Two)

5.1 Introduction.

Much has been said about the online experience, particularly in relation to general internet use and web-based interventions such as online CBT programs, online counselling, and peer-support forums. However, there has been relatively little empirical research directly examining the consumer experience of web-based clinical assessment programs in a naturalistic setting. While there is literature examining the acceptance and usability of computerised clinical assessment programs, their findings may not be relevant to the current online context, nor capture the subtle experiential aspects of a relatively distinct online program such as the e-PASS.

The views and reactions of consumers are considered important criteria of any diagnostic instrument (Pinninti et al., 2003), and this could be similarly applied to a web-based program producing diagnostic results. The e-PASS was designed to be a useful online diagnostic program in terms of being accessible, technically stable, easy to use, and informative, amongst other characteristics. Brief pilot testing indicated no major usability or experiential issues amongst research staff and a small sample of clients from a psychology clinic. However, further evaluation of the e-PASS experience was clearly needed, particularly with a sample of actual consumers completing the e-PASS under relatively naturalistic conditions. Hence, this study aimed to specifically investigate the user experience of the e-PASS.

Here, the term “user experience” is broadly applied to explore potentially diverse aspects of the construct in relation to the e-PASS. Facets of usability and user experience such as ease of use, affect, values, and so on were considered (discussed in more detail later). In exploring the user experience of the e-PASS, this study addressed a “summative usability” approach which, according to Tullis and Albert (2008) refers to two key criteria: whether usability goals of the “product” are met, and how well the “product” compares against its alternative (Tullis & Albert, 2008). In line with this, the usability/experience of the e-PASS was considered independently and in comparison with the Clinical Interview (CI; i.e., from Chapter Four) and other previously experienced forms of assessment. In other words, a focus of this study was how well the e-PASS performed in comparison with alternative clinical assessment instruments and methods.

While often recommended for user experience studies, the use of observation and talk-aloud protocols (Ericsson & Simon, 1993; Mehlenbacher, 2002) as data collection methods were decided against in the present study for several reasons. Firstly, having participants come into and perform within a controlled observational environment was considered too burdensome and

potentially distressing given the probable background of mental health issues. Furthermore, observation can result in a reactivity effect, where the authenticity of behaviour is compromised by the awareness of being observed (Krahn & Putnam, 2008). This could obscure the experience of one of the key characteristics of the e-PASS, which is the ability to perform it remotely in a private setting of one's choosing. Hence, instead of direct observation, this study evaluated the experience of e-PASS users via a questionnaire. While there are limitations to using questionnaires and retrospective data collection, this approach was viewed as advantageous in allowing respondents to reflect on their experiences after completing the e-PASS, and in comparison to their subsequent CI experience. It was presumed that important usability issues apparent in an observation/talk-aloud protocol would still emerge in retrospective data.

Specifically, data was collected via a multi-faceted online survey, created for this study, which asked participants who completed the e-PASS and the CI to rate their level of agreement with user experience-related statements (UE statements) referring to the e-PASS and CI. Using open questions, the survey also asked participants to compare the e-PASS to past assessments; and share their views of liked/disliked aspects of the e-PASS. Although there are several formal usability/user experience related questionnaires available, the decision to develop a new survey was based on the need for a measure that was more relevant to the likely issues of web-based assessment and the interests of this research. With sufficient data, a secondary goal was to perform factor analysis on the survey data to identify experiential factors.

5.1.1 Aims and hypotheses.

Given the little empirical evidence regarding online assessment programs, this study was largely explorative in investigating the experience of the e-PASS, with no a priori hypotheses. The central questions guiding this research were: what are the main qualities of the e-PASS, including the positive and negative characteristics perceived by users; how does the e-PASS experience differ from that of a structured clinical interview and previous assessment experiences; and does the e-PASS experience differ between different users. However, there were some general expectations based on the online literature. For example, as the online experience is often closely associated with convenience and anonymity (e.g., Barak & Hen, 2008), these themes were expected to apply with the e-PASS, particularly in contrast with more traditional clinical assessment methods. Also, given the lack of interviewer involvement and social judgment associated with this, the e-PASS was expected to elicit more self-disclosure and feelings of comfort than the CI in Study One and past experiences of traditional assessment.

Drawing on the clinical assessment (including the computerised assessment) literature, there are a range of experiential factors (miscomprehension, distraction, symptoms of psychopathology) that could compromise the validity of online clinical diagnostic programs. However, to the knowledge of the author, no studies to date have empirically examined this. Therefore, an additional study aim was to investigate to what extent UE factors contributed to the diagnostic accuracy of the e-PASS. This was performed by comparing UE statement ratings with diagnostic agreement outcomes (as previously determined in the previous study) to see if any experiential factors predicted diagnostic agreement between the e-PASS and the CI. Identifying significant predictors/correlates could help inform which experiential factors may need addressing in order to improve the diagnostic accuracy of the e-PASS.

5.2 Methodology.

5.2.1 Participants

All 158 participants who completed the CI from Study One were verbally offered at the end of their phone interview the opportunity to undertake the User Experience (UE) Survey. While 110 people agreed to this, 88 individuals ultimately accessed and completed the Survey (the Survey group). The mean age of the Survey group was 41.89 years (*SD* = 12.75) while the median age was 39.00 years. On average, the Survey group had 1.81 e-PASS clinical diagnoses (*SD* = 2.26), and 1.61 clinical diagnoses (*SD* = 1.56) from the CI. Other participant characteristics appear in the Results section of this chapter.

5.2.2 Materials.

The UE Survey was developed for this study with the purpose of measuring experiential themes elicited by the e-PASS, and in comparison to the CI of Study One and other forms of assessment previously experienced by respondents. The survey consisted of two parts: 1) self-rating statements and; 2) questions regarding previous assessment experience and perceived positive and negative aspects of the e-PASS.

1. *Self-Rating Statements*: This included 36 self-rated statements referring to various experiential themes (see Table 30) reported in the literature to be associated with the experience of online assessment programs and assessments in general. Themes included perceived anonymity (“I felt anonymous”), convenient accessibility (“It was convenient to access”), increased disclosure (“I felt like I could open up”), and motivation for follow-up (“Afterwards, I felt motivated to seek treatment”). Each of the

themes were represented by at least one statement ("It was convenient to undertake") and a reversed statement ("It was a hassle to access") added as an indicator of internal reliability. Statements were created by the student researcher and reviewed by supervisors and peers with the criteria of being easy to understand, distinct from one another, and relevant to the online experience. Statements were randomly ordered within the UE survey. Each statement was presented twice and consecutively, with one statement labelled to refer to the respondent's e-PASS experience, and the other labelled to refer to the respondent's CI experience. Hence, participants were effectively presented with 72 statements (i.e., two sets of 36 statements) to rate. Each statement used a nine-point Likert scale of agreement with labels for the end-points and intervals of two (0 = "not at all", 2 = "not really", 4 = "moderately", 6 = "highly", 8 = "definitely"). Preliminary testing of the survey suggested a completion time of between 10 to 15 minutes. However, actual completion times for participants in this study were not recorded.

2. *Questions regarding previous assessment experience and perceived positives/negatives of the e-PASS*: The second part of the survey comprised of 11 closed and open questions (see Table 31 below). Questions focused on the respondent's previous experience with online assessment programs and more traditional forms of assessments (e.g., consultation with GP or face-to-face diagnostic assessment by psychologist/psychiatrist), and how their experience of the e-PASS compared to these. There were also two questions regarding what the respondent liked and disliked about doing the e-PASS, and two questions asking participants to comment on what they perceived as being the advantages and disadvantages of doing a program such as the e-PASS. The survey concluded with an opportunity for respondents to state any further comments about their e-PASS experience.

Table 30
*Self-Rating Statements in the User Experience Survey**

	Intended theme
3. I would avoid this if I needed another assessment	Acceptance
8. I would do this again in future if I need another assessment	Acceptance
2. It was convenient to access	Accessibility
15. It was a hassle to access	Accessibility
7. It was thorough in exploring for symptoms	Assessment quality
9. It was narrow in exploring for symptoms	Assessment quality
1. I answered the questions without much care	Attitude
16. I took it seriously when answering the questions	Attitude
6. I felt comfortable while participating in it	Comfort
21. It was awkward to participate in	Comfort
4. It was engaging and interesting	Engagement
13. I found myself distracted	Engagement
19. It was boring and tedious	Engagement
25. I found it easy to concentrate	Engagement
5. I felt restricted in what I wanted to express	Expression
34. I felt like I could open up	Expression
22. I felt put off from seeking treatment after finishing	Impact
23. I felt worse after finishing	Impact
27. Afterwards, I felt better about myself	Impact
35. Afterwards, I felt motivated to seek treatment	Impact
10. Afterwards, I felt motivated to see a health professional to discuss my mental health	Intervention
12. It captured useful information about me	Quality
24. I trusted it was accurate	Quality
31. It was too irrelevant in the information it was collecting	Quality
33. I would not trust it to be accurate	Quality
18. The assessment did not feel confidential	Security
29. I felt anonymous	Security
32. The information I gave will be handled with confidentiality	Security
36. I did not feel anonymous	Security
30. I felt put off from seeking help after finishing	Treatment impact
11. It was difficult to perform	Usability
14. It was too long	Usability
17. I found it easy to understand	Usability
20. It was easy to perform	Usability
26. It was too brief	Usability
28. It was really hard to understand	Usability

Note. * Statements rated on a 0 (“not at all”) to 8 (“definitely”) Likert scale of agreement. The number at the beginning of each statement refers to the statement’s position in the order within the survey.

Table 31
Additional Questions in the User Experience Survey

Please enter the approximate number of times you have ever completed an online mental health assessment program (e.g. questionnaire, survey, test). If once or more, please state the program(s) and/or its website(s).
How does your experience of the e-PASS compare with that of the other online assessment program(s) you have completed?
Please enter the approximate number of times you have ever undertaken a mental health interview similar to the phone/face-to-face interview you participated in.
Please enter the approximate number of times you have previously undergone a mental health interview or assessment with each of the following: General Practitioner, Psychologist, Psychiatrist, Other.
How does your experience of the e-PASS compare with that of the mental health interview(s) you have previous participated in?
How does your experience of the interview for this research compare with that of the mental health interview(s) you have previously participated in?
What did you particularly *like* about doing the e-PASS?
What did you particularly *dislike* about doing the e-PASS?
What are the *advantages* of doing a program like the e-PASS?
What are the *disadvantages* of doing a program like the e-PASS?
If you have any more comments you would like to make about your e-PASS experience, please describe below.

5.2.3 Procedure.

Participants who expressed interest in undertaking the UE Survey were immediately sent an email with a survey link and instructions to complete the survey within the next few days. The survey was presented online via Opinio software with a unique web address and password protection. Participants were prompted to supply their email address (as previously provided in Study One) for identification. The survey commenced with a statement asking the respondent to

"think back to when you participated in the e-PASS and phone interview, and rate the following statements on how much you agree with them on a scale of 0 (not at all) to 8 (definitely)". Participants were then presented items from the UE Survey over several web-pages. All self-rating statements were compulsory while open questions were optional. At the end of the Survey, participants received a message thanking them for their participation in the study. Participants received no further contact as part of the study. Survey data was securely downloaded onto a password protected computer and analysed using Microsoft Excel and SPSS.

5.2.4 Data analysis.

Median rank differences in UE statement ratings.

To examine differences in ratings between and within sample groups, two non-parametric hypothesis tests were employed to address the non-normal distribution of UE statement data. Firstly, the Mann-Whitney *U* (M-W *U*) test was used to compare median differences of e-PASS statement ratings between groups within the sample. For example, the M-W *U* test was performed to examine whether individuals in the sample who expressed wanting a psychological assessment during the pre-e-PASS socio-demographic/background questions ("assessment" group) would rate the UE statements differently to those not wanting a psychological assessment ("non-assessment" group). The M-W *U* test is concerned with median differences and examines this by comparing the ranking distribution of scores between the two groups involved (Sheskin, 2000). A statistically significant M-W *U* test indicates there is a difference in the medians of the two groups. The test can be used with ordinal or interval data and is considered more statistically efficient (i.e. requires smaller sample to reach statistical significance) and robust in dealing with outliers than the student *t*-test.

The Wilcoxon Signed Rank (WSR) test is another non-parametric which is ideal for analysing median differences between two related samples derived from a population which cannot be assumed to be normal in distribution. The WSR test is the non-parametric equivalent of the paired *t*-test in that it deals with paired data from the same group (Field, 2009). In this study, the WSR test was used to test whether there were statistically significant differences between the agreement ratings of e-PASS and CI focused statements. The WSR test involves ranking the absolute difference between the ratings of paired statements (i.e. e-PASS and CI related statement), calculating the absolute value of the sum of the ranks which is referred to as the test statistic, *W*, then comparing it against the critical *W* score. If *W* is significant, the null hypothesis is rejected, and it can be concluded that the rating of paired statements is significantly different.

With a sufficiently large sample, a *z*-score can be calculated from both the WRS and M-W *U* statistics and be compared against a critical *Z* score to test statistical significance. The *Z* score can also be used to derive an effect size ("*R*") which can be interpreted against Cohen's criteria: .00 - .30 = weak; .30-.60 = moderate; greater than .60 = large (Field, 2009). By ranking the UE statement ratings rather than working with the data in its interval form, both the M-W *U* and WSR tests have little or no impact from outliers in the data (Sheskin, 2000). Therefore, the various outliers in the UE statement ratings were retained when performing the M-W *U* and WSR analyses.

Exploratory factor analysis of e-PASS related UE statements.

Exploratory factor analysis (EFA) was used to investigate broader user experience themes derived from the UE survey statement ratings. EFA is recommended for new datasets, where it is unclear how variables relate to one another, and whether underlying factors exist (Field, 2000). The aim of factor analysis is to reduce the dimensionality of a data set into factors which represent groups of variables that are highly correlated with one another, but have little correlation with variables outside of the factor (Field, 2000). The factor loading refers to the size of correlation between a variable and the factor it belongs to. The squared value of a factor loading can indicate the amount of variance a variable contributes to the factor, and therefore reflect how important the variable is for the factor. The factor score refers to the score of which a person has on a particular factor, and is based on the individual's scores on the variables that load onto the factor.

There are two main methods of EFA. Principal component analysis (PCA) assumes the total variance of the variables involved can account for all of the variance of the factors (Field, 2000). Conversely, factor analysis assumes error variance and is therefore considered more conservative than PCA. Factor anlaysis is also more dependent on the normality of variables involved, whilst PCA can still produce adequate dimensional representation with non-normal input variables. In general, PCA has less rigorous assumptions and requirements than factor analysis and was therefore used for EFA of the UE statements. A recent survey of over 1700 studies involving EFA found that over half of the studies used PCA (Costello & Osborne, 2005).

PCA involves a process of extracting principle components (factors), each with eigenvalues representing the amount of covariance between the variables within the factor. Factors are extracted in order of decreasing eigenvalues, with the first factor accounting for the most variance. The choice of which factors to retain is usually based on several "rules of thumb"

(Field, 2000): factors should have eigenvalues larger than one, a requirement referred to as the Kaiser criterion; factors should account for a total of 70-80% of the overall variance; and factors should only be retained if they are positioned above the “elbow” of the scree-plot (a plot of the factors and their eigenvalues; Costello & Osborne, 2005). It is also recommended that the communalities of the extracted factors are checked to ensure they represent a sizeable amount of the total variance.

Once decided which factors to retain, the factor loadings can then be rotated to assist in the interpretation of the factor structure. Of the two main types of rotation available, orthogonal rotation (varimax) is more commonly used and is simplest to interpret as it assumes the rotated factors are not correlated (Costello & Osborne, 2005). The alternative oblique rotation method was not used as it requires a priori assumptions of correlation between statements (Costello & Osborne, 2005) which was not the case in this study. Once rotated, the absolute values of factor loadings are examined to see which variables load onto the retained factors. As recommended by Tabachnick and Fidell (2001), a minimum loading of .32 was used to determine which statements were associated with a factor. A loading of .50 and above was considered strong, as recommended by Costello and Osborne (2005). Statements were excluded if they either: had insufficient loadings (i.e. below .32) on all retained factors; load (above .32) across two or more factors; or did not have two or more other statements loaded on the same component. Excluded statements were removed one at time, with the PCA repeated after each iteration until a factor solution was reached and fulfilled each of the above criteria for PCA.

One issue with PCA is determining an adequate sample size. A common rule of thumb is a sample size to item ratio of 10:1 (Field, 2000) while others have suggested that there should be at least 50 participants and at least five participants for every variable (Habing, 2003). However, some suggest the most important requirement is ensuring there are high factor loadings (Costello & Osborne, 2005); hence if there are relatively low loadings, then a larger sample is required. A practical check of whether a sample is sufficiently large is the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO-test). A KMO value above .50 indicates the sample size is adequate, whilst a value above .70 is considered good (Hutcheson & Sofroniou, 1999, as cited in Field, 2009). KMO values closer to 1 indicate that patterns of correlation are tight and factor analysis should yield distinct and reliable results (Hutcheson & Sofroniou, 1999, as cited in Field, 2009).

Qualitative analysis of UE Survey open questions.

Text responses to the UE survey open questions were collated into spreadsheets. A cross-case analytical technique was applied to examine the themes across survey respondents (Miles & Huberman, 1994). The method of constant comparison was used (Glaser & Strauss, 1967), which involves interpreting and coding statements, then comparing them with subsequent significant statements to allow consistency in coding. Afterwards, codes were grouped according to similarity, and assigned a thematic label based on the content of the groups. This process was reviewed by a peer and lead to minor changes in the coding of certain statements. The number of cases within a theme was then tallied up to reflect the proportion of the sample that expressed this theme. This method can indicate the quantity as well as the quality of the theme (Johnson et al., 2007), and is consistent with a quantitative approach to qualitative data (Morgan, 1993). The quantity and examples of relevant statements for each theme are presented in the Results section below.

5.3 Results.

5.3.1 Sample characteristics.

A total of 88 participants undertook the UE Survey after completing both the e-PASS and the CI. The Survey was declined by 70 participants, 15 of whom undertook the qualitative phone interview in Study Three as an alternative activity. The UE Survey was completed on average 13 days (min. - max. = 1 - 45 days, *SD* = 10.78) after completing the e-PASS. Socio-demographic variables of the Survey group were compared with the Total sample (*N* = 616) reported in Study One. The mean age (years) of the UE Survey group (*M* = 41.89, *SD* = 12.75) was significantly older than the remainder of the Total sample (*M* = 37.61, *SD* = 12.58; *t* = -2.95, *df* = 614, *p* < .01). Other demographic characteristics of the Survey group are depicted in Table 32 (below). Pearson's chi-squared testing showed no significant differences in the distribution of social demographic variables between the Survey group and Total sample, except in the area of post-schooling education (χ^2= 13.24, *df* = 5, *p* = .02). The UE group had a greater proportion of individuals who had completed an undergraduate or postgraduate degree, or a post-school (non-diploma) certificate.

Table 32

Comparing Demographic Variables of Total Sample and UE Survey Sub-sample

	Total sample % *N* = 616	UE survey group % *n* = 88	Chi-squared (*df*)
Gender			0.17 (1)
Male	28.1	26.1	
Female	71.9	73.9	
Relationship			6.96 (4)
Married	25.7	28.4	
Single	27.6	35.23	
De-factor	28.3	26.1	
Separated or divorced	10.8	9.5	
Other	5.6	4.9	
Country of birth			4.00 (5)
Australia	73.5	70.5	
United Kingdom	8.7	11.4	
Asian countries	5.1	4.6	
US	3.7	1.1	
European country (except UK)	3.7	5.7	
Other	6.0	3.4	
Setting			4.43 (3)
Metropolitan	62.3	65.9	
Regional	25.2	19.3	
Rural	10.6	10.2	
Remote	1.9	4.6	
Highest schooling			4.25 (3)
Year 9 or less	5.8	5.7	
Year 10	11.4	5.7	
Year 11	6.7	10.2	
Year 12	76.1	78.4	
Highest post-school education			13.24 (5) *
None	14.4	8.0	
Current undergraduate	13.4	4.6	
Undergraduate	23.4	28.4	
Postgraduate	19.0	22.7	
Diploma, apprenticeship, trade	14.9	13.6	
Certificate	14.8	22.7	
Employment			7.43 (5)
Full time/Part time	66.5	63.6	
Disability, maternity, sick leave	7.1	4.6	
Home duties/carer	7.0	4.6	
Retired	3.1	6.8	
Unemployed	10.2	11.4	
Other (e.g. volunteer, student)	6.0	9.1	

Note. All Chi-squared tests were statistically insignificant except * ($p < .05$)

As seen in Table 33, the same proportion of the UE Survey group (47%) were receiving current mental health assistance as seen in the Total sample. Across the individual services, the UE group had slightly higher rates of current access to a medical doctor, counsellor, self-help book, and slightly lower rates of access to a psychiatrist, mental health social worker, and mental health nurse; however, neither of these differences were statistically significant. Regarding current treatment, there were also no significant differences between the Survey group and the Total sample. However, there were higher usage rates of CBT and mood stabilisers, and lower usage rates of anxiolytics/hypnotics, amongst the Survey group compared with the Total sample.

Table 33
Comparing Mental Health Service and Treatment Use between Total Sample and UE Survey Group

	Total sample %	Survey group %	Chi-squared*
Receiving current mental health assistance	47.7	47.7	0.00
Current service type			
Medical doctor	26.0	30.0	0.58
Psychologist	21.1	22.7	0.14
Psychiatrist	13.6	11.4	0.39
Self-help book	10.1	13.6	1.24
Counsellor	5.5	8.0	1.00
Telephone service	2.1	2.3	0.01
Online service	2.1	2.3	0.01
Social worker	1.8	0.0	1.60
Mental health nurse	1.0	0.0	1.60
Current treatment type			
Antidepressants (e.g. Zoloft, Effexor)	25.8	24.0	0.17
General counselling	23.3	25.0	0.14
CBT	14.3	19.3	1.82
Anxiolytic or hypnotic (e.g. Valium)	8.3	4.6	1.60
Antipsychotic med. (e.g. Seroquel)	3.6	3.4	0.01
Mood stabilizer (e.g. Lithium)	2.8	5.7	2.80
Hypnosis	0.8	2.3	2.33

Note. Total sample $N = 616$, UE Survey group $n = 88$; * $df = 1$, with all values statically insignificant ($p > .05$).

5.3.2 User experience statements

Table 34 (below) presents the descriptive statistics of the 36 UE rating statements referring to the experience of the e-PASS and the CI. The range of ratings was between 0 (not at all) to 8 (definitely) for most statements, while the mean and median rating varied across the statements. An examination of histograms and skewness/kurtosis values indicated that most of the UE statements were not normally distributed. This was confirmed with a Shapiro-Wilk test of normality which was significant ($df = 88$, $p < .01$) for all 72 statements. Hence, the following analyses focused on median ratings and used non-parametric inference tests (e.g. Wilcoxon Rank Test) deemed more suitable for non-normally distributed data.

In general, there were high median ratings (6 and above) of agreement with statements about: perceiving the e-PASS as convenient; taking a serious approach to answering e-PASS questions; the e-PASS being easy to understand, confidential, and easy to perform. Conversely, participants strongly disagreed (0 to 2) with statements about answering e-PASS questions without care, finding the e-PASS too long, being distracted when doing the e-PASS, and being deterred from treatment. Several statements were rated with moderate levels of agreement (3 to 5). For example, participants on average moderately agreed with statements about feeling restricted in expression, “feeling better about myself”, feeling motivated to see a health professional for follow-up assessment, trusting the e-PASS to be accurate, and finding it easy to concentrate when doing the e-PASS. However, standard deviations showed greater variation in ratings for some items than others.

Table 34

Descriptive Statistics of the Agreement Ratings of e-PASS and Clinical Interview Related User Experience Statement Ratings

Statement	Theme	e-PASS			Clinical Interview		
		M	*SD*	Med.	*M*	*SD*	Med.
3. I would avoid this if I needed another assessment	Acceptance	1.16	1.91	0.00	1.13	1.92	0.00
8. I would do this again in future if I need another assessment	Acceptance	6.00	2.01	6.00	6.19	1.86	6.00
2. It was convenient to access	Accessibility	6.56	1.80	7.00	6.40	1.80	7.00
15. It was a hassle to access	Accessibility	1.32	1.87	1.00	1.05	1.70	0.00
1. I answered the questions without much care	Attitude	1.84	2.36	1.00	1.53	2.28	1.00
16. I took it seriously when answering the questions	Attitude	7.32	1.39	8.00	7.56	1.03	8.00
6. I felt comfortable while participating in it	Comfort	6.67	1.51	7.00	6.38	1.70	7.00
21. It was awkward to participate in	Comfort	1.08	1.64	0.00	1.53	1.86	1.00
4. It was engaging and interesting	Engagement	5.15	2.00	6.00	5.84	2.00	6.00
13. I found myself distracted	Engagement	1.72	2.16	1.00	1.27	1.92	0.50
19. It was boring and tedious	Engagement	1.90	2.02	1.00	1.43	1.97	1.00
25. I found it easy to concentrate	Engagement	5.09	2.24	5.00	5.67	2.15	6.00
5. I felt restricted in what I wanted to express	Expression	3.48	2.77	4.00	1.41	1.75	1.00
34. I felt like I could open up	Expression	5.14	2.59	6.00	6.16	2.12	7.00
22. I felt put off from seeking treatment after finishing	Impact	0.98	1.90	0.00	0.73	1.60	0.00
23. I felt worse after finishing	Impact	1.23	2.09	0.00	1.26	2.05	0.00
27. Afterwards, I felt better about myself	Impact	3.23	2.31	3.00	3.79	2.43	4.00
35. Afterwards, I felt motivated to seek treatment	Impact	3.47	2.73	4.00	3.91	2.67	4.00
10. Afterwards, I felt motivated to see a health professional to discuss my mental health	Intervention	3.51	2.66	4.00	3.94	2.65	4.00
7. It was thorough in exploring for symptoms	Quality	5.36	1.84	5.00	6.07	1.67	6.00
9. It was narrow in exploring for symptoms	Quality	2.60	2.16	2.00	1.90	1.97	1.00
12. It captured useful information about me	Quality	5.07	1.82	5.00	5.57	1.84	6.00
24. I trusted it was accurate	Quality	5.02	2.35	5.00	5.59	2.33	6.00
31. It was too irrelevant in the information it was collecting	Quality	1.41	1.70	1.00	1.08	1.69	0.00
33. I would not trust it to be accurate	Quality	1.82	2.14	1.00	1.12	1.63	1.00
18. The assessment did not feel confidential	Security	1.37	2.28	0.00	1.34	2.26	0.00
29. I felt anonymous	Security	5.40	2.53	6.00	4.44	2.65	5.00
32. The information I gave will be handled with confidentiality	Security	6.71	1.66	7.00	6.69	1.76	7.00
36. I did not feel anonymous	Security	1.29	1.88	1.00	2.09	2.45	1.00
30. I felt put off from seeking help after finishing	Treatment impact	0.75	1.31	0.00	0.63	1.20	0.00
11. It was difficult to perform	Usability	1.53	1.84	1.00	1.30	1.70	1.00
14. It was too long	Usability	2.07	2.30	1.00	1.73	2.14	1.00
17. I found it easy to understand	Usability	6.41	1.84	7.00	7.00	1.37	7.00
20. It was easy to perform	Usability	6.20	1.85	7.00	6.48	1.79	7.00
26. It was too brief	Usability	1.22	1.69	0.00	1.09	1.44	0.00
28. It was really hard to understand	Usability	1.06	1.55	0.00	0.69	1.37	0.00

Note. n = 88 * SE = 0.26, ** SE = 0.51; Med. = Median; Ratings on a 0 (“not at all”) to 8 (“definitely”) scale of agreement.

5.3.3 Differences in ratings of e-PASS and Clinical Interview related statements.

Of interest was whether participants perceived the e-PASS differently to the CI as reflected in their rating of UE statements. Given the non-normal distribution of most statement ratings, a Wilcoxon Ranked Signed (WRS) test was used to compare median differences in ratings of specific statements referring to the e-PASS and the CI. For example, the WRS test was used to indicate whether participants rated the item "It was convenient to access" significantly different for the e-PASS compared with the CI.

Despite similar median ratings between the e-PASS and CI related statements (Table 35), there were statistically significant differences across most pairs of statements, with the Wilcoxon effect size ranging from small to large in size. The largest rating difference was for the statement "I felt restricted in what I wanted to express", with a median rating of 4 (moderate agreement) for the e-PASS and 1 (Not at all) for the CI ($Z = -5.99$, $p < .001$). There were also other large differences indicating that participants perceived the CI as: easier to understand, more thorough, less narrow in exploring for symptoms, and less boring and tedious than the e-PASS.

To a lesser extent, participants also rated the e-PASS as longer, yet capturing less useful information than the CI. Other moderate differences in ratings also indicated participants were less engaged with the e-PASS and answered the questions with less care than when performing the CI. However, participants rated feeling more anonymous and less awkward when participate in the e-PASS compared with the CI. There were a few non-significant differences in ratings which suggested the e-PASS and CI were similarly perceived in terms of level of convenience, confidentiality, and difficulty to perform. Lastly, participants on average strongly disagreed when rating both the e-PASS and CI related statements about feeling worse after completing the assessment, whilst moderately agreeing (rating of 6) that they would do both the e-PASS and CI again in future if they needed another assessment.

Table 35

Wilcoxon Ranked Signed Test Results Comparing Median Difference of Ratings Between e-PASS and Clinical Interview Statements

Statement	Theme	e-PASS Med.	CI Med.	*Z*	Effect size	
3. I would avoid this if I needed another assessment	Acceptance	0.00	0.00	-	-	-
8. I would do this again in future if I need another assessment	Acceptance	6.00	6.00	-	-	-
2. It was convenient to access	Accessibility	7.00	7.00	-	-	-
15. It was a hassle to access	Accessibility	1.00	0.00	-		
1. I answered the questions without much care	Attitude	1.00	1.00	3.04**	.32	Medium
16. I took it seriously when answering the questions	Attitude	8.00	8.00	-2.57*	-.27	Medium
6. I felt comfortable while participating in it	Comfort	7.00	7.00	-2.59*	-.28	Small
21. It was awkward to participate in	Comfort	0.00	1.00	-3.41**	-.36	Medium
4. It was engaging and interesting	Engagement	6.00	6.00	3.56***	.38	Medium
13. I found myself distracted	Engagement	1.00	0.50	-2.59*	-.28	Small
19. It was boring and tedious	Engagement	1.00	1.00	-3.20**	-.34	Medium
25. I found it easy to concentrate	Engagement	5.00	6.00	-3.44**	-.37	Medium
5. I felt restricted in what I wanted to express	Expression	4.00	1.00	-5.99***	-.64	Large
34. I felt like I could open up	Expression	6.00	7.00	-3.61***	-.38	Medium
22. I felt put off from seeking treatment after finishing	Impact	0.00	0.00	-2.59*	-.28	Small
23. I felt worse after finishing	Impact	0.00	0.00	-	-	-
27. Afterwards, I felt better about myself	Impact	3.00	4.00	-3.47**	-.37	Medium
35. Afterwards, I felt motivated to seek treatment	Impact	4.00	4.00	-2.63**	-.28	Small
10. Afterwards, I felt motivated to see a health professional to discuss my mental health	Intervention	4.00	4.00	-2.75***	-.29	Medium
7. It was thorough in exploring for symptoms	Quality	5.00	6.00	-4.48***	-.48	Large
9. It was narrow in exploring for symptoms	Quality	2.00	1.00	-4.15***	-.44	Large
12. It captured useful information about me	Quality	5.00	6.00	-3.62***	-.39	Medium
24. I trusted it was accurate	Quality	5.00	6.00	-3.29**	-.35	Medium
31. It was too irrelevant in the information it was collecting	Quality	1.00	0.00	-3.32**	-.35	Medium
33. I would not trust it to be accurate	Quality	1.00	1.00	-4.04***	-.43	Medium
18. The assessment did not feel confidential	Security	0.00	0.00	-	-	-
29. I felt anonymous	Security	6.00	5.00	-4.35***	-.46	Medium
32. The information I gave will be handled with confidentiality	Security	7.00	7.00	-		
36. I did not feel anonymous	Security	1.00	1.00	-4.15***	-.44	Medium
30. I felt put off from seeking help after finishing	Treatment impact	0.00	0.00	-2.00*	-.21	Small
11. It was difficult to perform	Usability	1.00	1.00	-	-	-
14. It was too long	Usability	1.00	1.00	-2.96**	-.32	Medium
7. I found it easy to understand	Usability	7.00	7.00	-3.40**	-.36	Medium
20. It was easy to perform	Usability	7.00	7.00	-2.23*	-.24	Small
26. It was too brief	Usability	0.00	0.00	-	-	-
28. It was really hard to understand	Usability	0.00	0.00	-3.42**	-.36	Medium

Note. $n = 88$; * $p < .05$; ** $p < .01$; *** $p < .001$; Med. = median; ratings on a 0 ("not at all") to 8 ("definitely") scale of agreement.

5.3.4 Differences in ratings within sample groups.

A Mann-Whitney *U* (M-W *U*) test was used to compare median differences of e-PASS statement ratings between groups within the sample (e.g. men vs. women). For example, the M-W *U* test was performed to examine individuals in the sample who expressed wanting a psychological assessment during the pre-e-PASS survey ("assessment" group) would rate the statements differently to those not wanting a psychological assessment ("non-assessment" group). As seen in Table 36, the assessment group rated higher agreement on statements indicating they would less likely avoid the e-PASS if they needed another assessment. Median differences also suggested the assessment group felt more motivated to see a health professional following the e-PASS, being less distracted during the e-PASS, answering e-PASS questions more seriously, and finding it briefer than the non-assessment group. The median differences were all considered small in size.

Table 36
Mann-Whitney U Test Comparing Median Differences of Ratings Between Assessment and Non-Assessment Seekers in Sample

Statement	Assessment *n* = 46	Non-assessment *n* = 41	M-W *U*	*Z* Value*	Effect size
3. I would avoid this if I needed another assessment	39.01	49.05	735.50	-1.98	Small
10. Afterwards, I felt motivated to see a health professional to discuss my mental health	49.57	38.95	733.00	-1.96	Small
13. I found myself distracted	39.41	50.07	732.00	-2.04	Small
16. I took it seriously when answering the question	49.07	39.50	756.00	-2.24	Small
25. I found it easy to concentrate	51.33	37.02	652.00	-2.65	Small
26. It was too brief	48.63	38.08	730.00	-1.96	Small

Note. All *Z* values significant at $p < .05$

Gender also had a significant effect on the rating of two e-PASS related statements. Specifically, men rated higher agreement than women for the statements "it captured useful information about me" (M-W $U = 500.00$, $Z = -2.39$, $p = .017$), and "I felt like I could open up" (M-W $U = 407.00$, $Z = -2.74$, $p = .006$). When participants were grouped by service access, participants currently accessing one or more services (e.g., GP, psychologist, psychiatrist) more

highly agreed with the statements “it was convenient to access” (M-W U = 527.00, Z = -2.09, p = .037) and “afterwards, I felt motivated to see a health professional to discuss my mental health” (M-W U = 597.00, Z = -2.61, p = .009) than non-current service users.

Median differences in e-PASS related statement ratings were also compared between individuals with and without a clinical diagnosis according to the CI. Individuals with social phobia rated their agreement with the statement “It was awkward to participate in” significantly higher than individuals without social phobia (M-W U = 627.00, Z = -1.98, p = .048). Participants with GAD had a significantly lower median agreement rating with the e-PASS related statements “It captured useful information about me” (M-W U = 441.00, Z = -2.35, p = .019) and “I would not trust it to be accurate” (M-W U = 417.00, Z = -2.42, p = .015). Compared with participants without MDD, participants with a clinical diagnosis rated higher median agreement with several statements: “It was narrow in exploring symptoms” (M-W U = 417.00, Z = -2.22, p = .026), “It was too long” (M-W U = 614.50, Z = -2.40, p = .016), “It was too irrelevant in the information it collected” (M-W U = 613.50, Z = -2.03, p = .043), and “I felt restricted in what I wanted to express” (M-W U = 593.50, Z = -2.55, p = .011).

There were also median differences on the statement “Afterwards I felt motivated to see a health professional to discuss my mental health” based on the number of e-PASS subclinical/clinical diagnoses (M-W U = 688.50, Z = -2.18, p = .030). Those with five or less diagnoses rated lower agreement on this statement than those with six or more e-PASS diagnoses.

The Kruskal-Wallis H test (a variant of the M-W U test) was used to analyse differences in statement ratings for variables with more than two categories. From this, quality of life ratings resulted in significant differences in particular statement ratings. Generally, lower quality of life ratings associated with higher agreement for the statements “I felt restricted in what I wanted to expression” (χ^2 = 10.43, df = 4, p =.034) and “It was boring and tedious” (χ^2 = 13.36, df = 4, p =.010), and lower agreement for the statements “I found it easy to concentrate” (χ^2 = 15.66, df = 4, p =.004) and “Afterwards I felt better about myself” (χ^2 =11.59, df = 4, p =.021). In other words, participants reporting lower quality of life appeared to experience relatively more restricted expression, more feelings of boredom, more difficulties concentrating, and less self-esteem after the e-PASS.

Learning styles also associated with differences in two statement ratings. Those who preferred to learn by reading generally had higher agreement with the statement “I trusted it was accurate” (χ^2 = 8.45, df = 3, p =.038) and “afterwards, I felt better about myself” (χ^2 = 10.03, df =

3, p =.018), compared with those who preferred to learn by hearing and looking/watching. There were no significant median differences in statement ratings between residential settings (i.e. metropolitan, regional, rural), though ratings for the statement “it was a hassle to access” approached significance. Education level also did not associate with significant rating differences.

5.3.5 User Experience and its relationship with diagnostic agreement between the e-PASS and the Clinical Interview.

Of further interest was whether user experience variables related to the diagnostic agreement between the e-PASS and the CI. Diagnostic agreement referred to when a diagnostic result (i.e., clinical diagnosis or non-clinical diagnosis) for a particular disorder was given by both the e-PASS and the CI. Due to the small sample size, analysis focused on disorders with at least 20% diagnostic disagreement, in order to compare reasonable participant group sizes (i.e., of those with and without diagnostic agreement). The disorders that met this criterion were MDD, social phobia, insomnia, and GAD. A Mann-Whitney *U* test was used to compare median differences in statement ratings between those with and without diagnostic agreement. As seen in Table 37, diagnostic agreement had a small to moderate effect size on the median rating of specific statements.

Table 37

Mann-Whitney U Test Comparing Median Differences of e-PASS Experience Statement Ratings Between Participants with and without Diagnostic Agreement for Specific Disorders

e-PASS Experience Statement	Median Rank		M-W *U*	*Z* value	Effect size
	Agree.	Non-agree.			
MDD	*n* = 69	*n* = 18			
13. I found myself distracted	41.01	55.47	414.50	-2.25*	Small
14. It was too long	39.85	59.92	334.50	-3.08**	Mod.
30. I felt put off from seeking help after finishing	41.60	51.21	418.00	-2.52*	Small
Social phobia	*n* = 70	*n* = 18			
1. I answered the questions without much care	41.49	54.35	419.00	-1.98*	Small
21. It was awkward to participate in	40.76	57.35	368.00	-2.66**	Small
GAD	*n* = 64	*n* = 23			
19. It was boring and tedious	40.12	54.80	487.50	-2.46*	Small
14. It was too long	40.12	54.80	487.50	-2.46*	Small
13. I found myself distracted	40.97	52.43	542.00	-1.94	Small
Insomnia	*n* = 70	*n* = 17			
10. Afterwards, I felt motivated to see a health professional to discuss my mental health	40.58	58.09	355.50	-2.59*	Small

Note. * $p < .05$; ** $p < .01$; M-W *U* = Mann-Whitney *U* test statistic.

Participants with diagnostic disagreement for MDD had significantly higher levels of agreement with statements referring to: being distracted during the e-PASS; perceiving the e-PASS as too long; and feeling put off from seeking help after finishing. Participants with diagnostic disagreement for social phobia appeared to agree more with statements about feeling awkward and completing the e-PASS "without much care" than individuals with diagnostic agreement. Diagnostic disagreement for GAD also resulted in significantly higher ratings of agreement with statements about the e-PASS being boring and tedious, too long, and feeling distracted. Lastly, individuals with diagnostic disagreement for insomnia had higher agreement ratings to the statement referring to the need to discuss their mental health with a health professional.

Logistic regression was also performed to see if UE statement ratings referring to the e-PASS predicted diagnostic agreement/disagreement between the e-PASS and CI. Again, only

disorders with at least 20% diagnostic disagreement were considered. Table 38 displays the regression model properties of the statements which significantly predicted agreement.

Table 38
Binary Logistic Regression of Diagnostic Agreement using e-PASS Experience Statement Ratings as Predictors

	Model χ^2 (*df* = 1)	-2LL	H-L test (*p*)	*B*	*SE* (*B*)	Wald's χ^2 (*B*) (*df* = 1)	Odds ratio (e^B) (95% CI)	AUC (95% CI)
MDD								
"It was too long"	9.95	78.76	.87	-0.35	0.11	9.35**	.71 (.57- .83)	.73 (.60-0.86)
Social phobia								
"It was awkward to perform"	4.99	78.04	.08	-0.42	0.15	7.54**	0.66 (.49-.89)	.69 (.53-.85)
Insomnia								
"Afterwards, I felt motivated to see a health professional"	7.88	79.17	.25	-0.28	0.11	6.21*	0.76 (.61-.94)	.70 (.57-.83)

Note. * $p < .05$ ** $p < .01$; AUC = Area Under the Curve; CI = confidence interval; H-L = Hosmer-Lemeshow; LL = log likelihood statistic.

Consistent with the Mann-Whitney *U* test results, higher agreement rating of the statement "It was too long" increased the likelihood of diagnostic disagreement for MDD, while higher agreement rating of the statement "it was awkward to perform" increased the likelihood of SAD diagnostic disagreement. Interestingly, higher agreement rating of the statement "afterwards I felt motivated....health professional" predicted insomnia diagnostic disagreement. While all coefficient variables were significant, the low odds ratios indicated a small change in the odds of diagnostic disagreement with a one unit increase in agreement rating. AUC values indicated low to medium goodness-of-fit of the regression models, suggesting that these UE statement ratings alone did not adequately explain diagnostic discrepancy.

5.3.6 Differences in User Experience ratings of e-PASS and CI as a predictor of diagnostic agreement.

Binary logistic regression modeling was then performed to examine whether differences between UE statement ratings for the e-PASS and CI associated with diagnostic disagreement. The differences between agreement ratings of corresponding e-PASS and CI related statements were created as new variables and entered as predictors. Only statements with medium to large

effects sizes (as previously identified in Table 35) were considered. For example, assessment type (e-PASS vs. CI) was previously found to have a large effect on the rating of the UE statement “It was narrow in exploring for symptoms”. Therefore, participants’ rating of this statement in relation to the CI was subtracted from their rating of the statement in relation to the e-PASS, to form a new variable which was inserted as a predictor for regression analysis. Logistic regression was performed on the diagnostic agreement results for disorders with at least 20% disagreement or agreement: MDD, panic disorder, social phobia, and GAD. All predictors were entered simultaneously and a Backward LR method of logistic regression was applied. The results are presented in Table 39.

Table 39

Binary Logistic Regression of Diagnostic Agreement using Differences between e-PASS and CI User Experience Statement Ratings as Predictors

	B (*SE*)	Wald's χ^2 (*B*)[(a)]	Odds ratio (e^B) (95% C.I.)	Model χ^2 [(a)]	-2LL	H-L test (*p*)	AUC (95% CI)
MDD							
"It captured useful information about me"	0.67 (0.26)	6.99**	1.96 (1.19 - 3.23)	12.77	75.93	3.68	.72 (.58 – .85)
"It was boring and tedious"	0.55 (0.28)	4.01*	1.73 (1.01 - 2.97)				
Social phobia							
"I answered without much care"	-0.53 (0.26)	4.03*	0.59 (0.35 - 0.99)	7.82	78.13	6.12	.72 (.58 - .86)
"It was awkward to participate in"	-0.46 (0.23)	3.89*	0.63 (0.40 - 1.00)				
GAD							
"I answered without much care"	-0.57 (0.25)	5.33**	0.57 (0.35 - 0.92)	6.16	94.34	0.74	.61 (.47 – .75)
Insomnia							
"It was engaging and interesting"	-0.58 (0.27)	4.62*	0.56 (0.33 - 0.95)	19.58	65.93	8.33	.79 (.67 - .91)
"It was thorough in exploring for symptoms"	0.88 (0.32)	7.65**	2.42 (1.29 - 4.52)				
"It was awkward to participate in"	0.57 (0.27)	4.39*	1.78 (1.04 – 3.04)				
"I felt anonymous"	0.52 (0.22)	5.64*	1.68 (1.10 - 2.58)				

Note. * $p < .05$ ** $p < .01$; (a) $df = 1$; AUC = Area Under the Curve; CI = confidence interval; H-L = Hosmer-Lemeshow; LL = log likelihood statistic.

Regarding the diagnosis of MDD, individuals with greater disparity between ratings of the e-PASS focused and CI focused statements "It captured useful information about me" and "It was boring and tedious" were more likely to have diagnostic disagreement. This suggests that participants who perceived the e-PASS as more relevant and/or boring and tedious than the CI have more likelihood of receiving different diagnostic results for MDD between the e-PASS and CI. The two predictor model resulted in an AUC (.72) indicating moderate overall accuracy in predicting diagnostic agreement. Differences in the UE ratings of two statements ("I answered without much care", "It was awkward to participate in") significantly predicted diagnostic

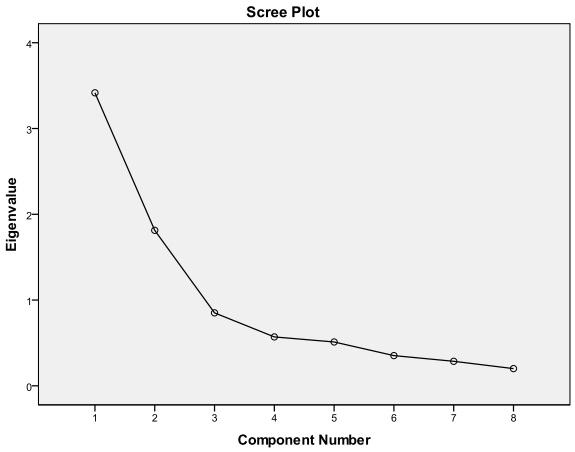
disagreement for social phobia, though the low odds ratio for both statements suggested they contributed little to prediction.

A one predictor model resulted for the regression of GAD diagnostic agreement. However, as reflected in the low odds ratio for the predictor and fairly low AUC, this model did not do well in accurately predicting agreement. The regression modeling of diagnostic agreement for Insomnia led to a four predictor model. Out of the four associated statements, the difference in rating of the statement "It was thorough in exploring for symptoms" produced the highest amount of contribution to prediction, while the statement "It was engaging and interesting" was associated with the lowest amount of contribution according to the odds ratio. The four predictor model resulted in an AUC value that indicated a medium level of accuracy in predicting diagnostic disagreement. Finally, the regression of diagnostic agreement for panic disorder resulted in no significant predictors in the regression model.

5.3.7 Exploratory factor analysis.

It was apparent that factor analysis of the UE statements data would unlikely produce stable results given the small sample size (n = 85). However, a principal component analysis (PCA) was performed on the 36 UE statements referring to the e-PASS to provide a preliminary investigation into the potential factors underlying the UE data. An iterative process was implemented involving the removal of statements displaying unsatisfactory communality or factor loadings (e.g. high loadings on two or more components), and repeating the PCA until an acceptable set of components and statements was formed. The process resulted in a two component model with eight statements. The main tests supported the retention of the eight statements to be used in component extraction: KMO measure of sampling adequacy for the overall set of eight statements was .72 (i.e. above the minimum requirement of .50); measures of sampling adequacy for each of the eight statements were also above .50; Bartlett's test of Sphericity was significant (approximate χ^2 = 279.59, df = 28, p <.001); and communality values indicated that at least 50% of the variance in each of the eight statements would be accounted for by potentially extracted components. The summary of explained variance showed the two resulting components had large eigenvalues (1.81 to 3.42), and together accounted for 65.37% of the variance. The scree plot for the eight items (see Figure 7) appeared to have an elbow or bend at component three, with a gradual decline in eigenvalues for subsequent components. Following the guideline of accepting components to the left of the elbow, the scree plot also supported keeping two components.

Figure 7. Scree Plot of Eight Components



A bivariate correlation was performed on the component scores and showed that components were clearly not correlated ($p < .001$), therefore justifying the use of orthogonal (varimax) rotation. The rotated component loadings of the eight statements on the two components are depicted in Table 40. As can be seen in the table, each statement solely loads onto one of the two components, and each component has four statements loaded onto it.

Table 40
Rotated Component Loadings of the UE e-PASS Statements

Statement	Component	
	1	2
9. It was narrow in exploring for symptoms	.88	
7. It was thorough in exploring for symptoms	-.80	
31. It was too irrelevant in the information it was collecting	.76	
5.I felt restricted in what I wanted to express	.73	
21. It was awkward to participate in		.83
6. I felt comfortable while participating in it		-.80
22. I felt put off from seeking treatment after finishing		.80
30. I felt put off from seeing a health professional after finishing		.74

Note. n = 85. Absent cells refer to component loadings below .25

The scores of two statements with negative loadings were reversed to have the same directionality as the remaining six statements. The Cronbach's alpha was found to be acceptable (i.e., above .60) for component one (.80) and two (.81), indicating both components displayed adequate internal consistency and did not require the addition or removal of statements. Component one appeared to indicate a negative theme around the perceived lack of information collected by the e-PASS. The statement with the highest rotated loading referred to the view that the e-PASS was narrow in its breadth of symptoms. A theme of discomfort and discouragement

for follow-up seemed to underlie component two. The statement referring to feeling awkward while participating in the e-PASS had the highest rotated factor loadings.

Using a standard regression method, the factor scores of the two components were calculated. Several outliers were detected and removed, leaving the distribution of the two components generally normal in characteristics. A comparison of means found no significant differences in regression factor scores based on: gender; residential setting; relationship status; whether or not a person is currently seeking professional assistance; whether or not a person saw a health professional in the past 12 months; and quality of life. Only one diagnosis related grouping was found to result in significant differences in factor scores. Specifically, people with diagnostic disagreement for social phobia had significantly higher mean scores on the discomfort and discouragement factor (factor 2) than those without diagnostic disagreement for social phobia (F [1, 79] = 6.67, p = .012).

5.3.8 UE Survey (open questions).

Previous online clinical assessment experience (both traditional and online).

Participants reported completing an "online mental health assessment program" (i.e. questionnaire, survey, test) similar to the e-PASS a mean average of 0.45 (SD = 0.93) times. However, almost 75% of e-PASS users indicated they had not completed such a program, and those who did referred to formal or established mental health websites such as *Moodgym*, *beyondblue*, *eCouch*, and *Beacon*. Twenty three participants commented about how e-PASS compared with other online programs. Most comments were in favor of the e-PASS, describing it as more thorough or comprehensive (30%), "personal", easier to use, and more informative. Only two participants described the e-PASS as similar, while four participants referred to the e-PASS as being less engaging/entertaining and more confusing than previously completed programs (e.g., measures at Moodgym).

Experience of e-PASS compared with previous assessment interviews.

In response to the question of how the experience of the e-PASS compared with that of previous assessment interviews (e.g., with a GP or psychologist), approximately 30% of the sample left this section blank. However, notably, almost a quarter of the sample (23.9%) described the e-PASS as more clinically comprehensive and/or in depth than previously undertaken assessments, while only 3.4% of participants referred to the opposite. More participants specified a preference of undertaking an interview with a health professional (e.g., "I

prefer to talk to someone") than doing the e-PASS (9%), as opposed to preferring the e-PASS over an interview (3.4%), while approximately 8.0% referred to the two as similar. The e-PASS experience was also described in various positive terms, such as being more comfortable to perform (4.6%), encouraging of self-disclosure ("...easier to talk to a computer than face to face"; 2.3%), and reassuring (1%) than previous assessment interviews. However, the e-PASS was also described as less personal ("...felt unconnected"; 4.6%), accurate (3.4%) and flexible ("I could not go back and change my answers"; 1.1%).

Experience of CI compared with previous assessment.

The CI was generally described with positive references and in many cases, similar to what people had previously experienced. A number of people described the CI as more thorough, structured, and in-depth than previous interviews with clinicians (e.g. GP, psychiatrist), which some participants found more satisfying and reassuring. Others however found it was too focused on "information collection", impersonal, and rigid in structure. Interestingly, participants referred to the phone medium as offering a sense of anonymity, but lacking the interpersonal connection associated with face-to-face interviewing and general counseling/therapy context. However, comments about CI interviewers were mostly positive (e.g. friendly, courteous) with infrequent minor criticisms (e.g. "interviewer seemed a bit timid…").

"Liked" aspects of the e-PASS experience.

This question was not responded to by approximately 28% of the sample. The remaining 63 responses were coded into experiential themes as presented in Table 41 below. The most commonly reported themes referred to convenience of access (18.2%), receiving assessment and diagnostic feedback (15.9%) and experiencing a sense of privacy (13.6%). Participants referred to accessibility by stating they could complete the e-PASS at a low cost, at a time of their choosing (e.g. "do not need to take time off work to make appointments) and in a comfortable and private setting. Regarding e-PASS feedback, some participants mentioned that they appreciated receiving a result, and that results were "objective", "reassuring", and a reflection of what they disclosed during the e-PASS. Participants who mentioned ease of use (10.2%) described the e-PASS as uncomplicated, "clear cut", and easy to respond to the multiple choice questions. Several participants (6.82%) highlighted the benefit of having autonomy over the e-PASS with regards to being able to complete it at their own pace, "let out as much as I wanted without being interrupted", and "without feeling time pressure". Of those who listed anonymity, privacy or

confidentiality as a liked aspect, only one participant extended on this, stating that they found it reassuring they could complete the e-PASS with “anonymity when feeling vulnerable”.

The e-PASS experience was also described as a means of enhancing insight (“it helped pin point specifics about myself”; 5.7%) and prompting follow-up evaluation. As one person stated, “the questions made me think about my condition, and…make me think that I should…see a GP who specialises in mood disorders to a get a professional opinion, which I then did”. A small number of participants specifically described the e-PASS as informative (3.4%; “I was able to know more about OCD”). The e-PASS was also highlighted as accurate and trustworthy (4.6%) because of its “objectivity” and association with a university. There were a few references to the e-PASS facilitating self-disclosure through its questionnaire format (“hard to express my feelings, but put in a question makes it much easier”) and privacy (“lack of personal scrutiny”).

Table 41
Frequency and Percentage of References to Experiential Themes in UE Sample

	n	%
Liked aspects of the e-PASS		
Blank	25	28.4
Accessibility/convenience	16	18.2
Assessment, diagnostic feedback	14	15.9
Privacy, anonymity, confidentiality, non-judgemental	12	13.6
Ease of use	9	10.2
Autonomy	6	6.8
Comprehensiveness, breadth	5	5.7
Insight, reflection	5	5.7
Perceived accuracy (e.g. valid, trustworthy)	4	4.6
Informative	3	3.4
Disliked aspects of the e-PASS		
Blank	32	36.4
None	6	6.8
Inflexible response	16	18.2
Length	9	10.2
Negative affect (e.g. bored, confronting, distressing)	7	8.0
Lack of sophistication	6	6.8
Repetitive	6	6.8
Impersonal	5	5.7
Misinterpretation (e.g. inaccurate)	5	5.7
Incomprehension (e.g. too wordy, confusing)	3	3.4
Crash	2	2.3

Note. $n = 88$

Disliked aspects of the e-PASS experience.

Thirty six percent of the sample (32 out of 88) left this section blank while several participants stated they did not have any dislikes of the e-PASS experience. Of the various negative experiential factors identified, the most commonly reported (18.2%) referred to difficulties in responding to the e-PASS. Some participants stated this in terms of disliking the multiple choice aspect of responding, and not being able to clarify their responses, for example via open responses. Others mentioned how they did not entirely fit into the categorical response options (e.g., yes/no) and therefore misrepresented their symptoms by having to select an option to progress in the program. There was a suggestion of including a "maybe" option to help respond to such questions. Amongst those who felt the e-PASS constrained expression, there were some who recognised this as a limitation of online surveys in general, and were therefore more accepting of it.

Some participants also seemed to particularly dislike the lengthiness (10%) of the e-PASS and the perceived repetition of the questions (6%) with some relating the two (e.g. "...long and I felt the questions were too similar"). A few people recommended including an option to save and return to the e-PASS and to display a progress indicator for users to know how long they have to go. Several participants found the e-PASS confusing (2%), overly wordy (2%), or too generalised (2%).

There were several references to the experience of negative affect. Some people reported finding the e-PASS "boring" and losing motivation due to the perceived lengthiness and repetition of the e-PASS. The process of reflecting on symptoms appeared to elicit discomfort for some individuals. For example, one participant said they particularly disliked the e-PASS because it "brought up old distresses and tensions" while another participant, perhaps experiencing a more subtle negative effect, said they found it "hard to focus" after completing the e-PASS. There were a few reports of people finding the warnings of suicide risk during the e-PASS alarming. As one participant stated "the epass (sic) warnings made me feel scared...as if I was so different that I really needed immediate help". Two participants were concerned that their "thoughts of death and dying" were misconstrued as "suicidal tendencies", when they in fact reflected fears of dying.

Only two participants raised the lack of support or follow-up as a particular issue while three separate individuals referred to the e-PASS as impersonal. One individual, who earlier listed anonymity as a positive aspect of the e-PASS, also referred to it here as a weakness ("anonymity means less face to face human interaction and little to no therapeutic alliance"). Almost 6% of the sample referred the e-PASS's lack of sophistication in distinguishing important causal and

maintaining factors behind their mental health problems. For example, one individual stated that the “e-PASS didn’t link my sleeping problems to my anxiety issues” whilst another individual was disappointed that the “questions didn't distinguish between situational and internal” factors that were contributing to their problems.

Table 42

Frequency and Percentage of Perceived Advantages and Disadvantages of the e-PASS

Theme	Example verbatim responses	*n*	%
Advantages			
Blank		17	19.3
None		3	3.4
Convenience	“able to be done at home in my own time”	21	23.9
Anonymous	“it feels anonymous”	14	15.9
Autonomy	“could take your time thinking about the answers”	10	11.4
Lack of human contact	“didn't have to set up communication with a stranger”	10	11.4
Privacy/confidentiality	“it felt totally private”	9	10.2
Comfort	“I could do it in the comfort of my own home”	8	9.1
Ease of access	“easy to access”	8	9.1
Insight/self-reflection	“made me think of my symptoms in more detail”	7	7.9
Low cost	“able to get an assessment without having to pay”	6	6.8
Self-disclosure	“I felt I could be completely honest”	5	5.7
Assessment	“It is a quick check of the state of one's mental health”	4	4.6
Understanding	“The questions, was like yes that's me!”	3	3.4
Encouragement	“proactive” “motivating”	2	2.3
Reputable/trustworthy	“created and run by reputable organisation”	2	2.3
Disadvantages			
Blank		30	34.1
None		9	10.2
Response limitations	“did not allow answers that needed a ‘depends…’”	15	17.0
Impersonal	“don’t have any personal contact”	12	13.6
Superficial/generic	“It feels robotic”	10	11.4
Unclear questions	“not enough info to indicate what is meant by question”	7	7.9
Time consuming	“it felt really long and tedious”	6	6.8
Inaccurate	“Not as precise as a face-to-face interview”	5	5.7
Disengaging	“Very easy to get distracted”	4	4.6
Technical problems	“My internet connection timed out a few times”	3	3.4

Note. n = 85

Underlying some participants’ reports was a relationship between several disliked factors, such as the inability to revise and clarify responses, and the e-PASS’s misinterpretation of

responses (5.7%). For example, one participant noted that the e-PASS misattributed their prescribed medication use as part of a substance disorder, but they could not go back to modify their earlier responses to avoid further substance abuse related questions and feedback. Another participant with a similar experience expressed this issue as relating to the incorrect assumptions the e-PASS made regarding substances. Only two participants (2%) reported a “crash” in the program and both said it happened after trying to go back and review/revise previous responses. In one case, the e-PASS was resumed after re-entering the website.

Perceived advantages and disadvantages of doing the e-PASS.

Table 42 also presents the categories of reported advantages of doing the e-PASS, including the frequency of references amongst the sample. The most frequently stated advantages were of convenience and anonymity. A number of participants also appreciated the ability to complete the e-PASS independently (Autonomy, 11.4%) which allowed participants to have more time to consider and respond to e-PASS items and “not be interrupted”. A few participants described the lack of human contact as advantageous because it alleviated pressure to respond a certain way, allowed them to be more honest, and circumvented the hassle of arranging appointments. Approximately eight percent recognised the e-PASS’s ability to raise insight or awareness of their mental health symptoms as an advantage while six participants acknowledged the low cost of the e-PASS as an advantage.

Forty four percent of the sample either stated there were no disadvantages of doing the e-PASS or left this section blank. The remaining participants reported at least one issue. The most frequently stated disadvantage related to constrained self-expression and difficulties with responding to items. Participants wrote about not being able to include contributing factors when responding to questions. For instance, one individual noted having major physical issues which influenced her mental health symptoms, but not being able to explain this via the e-PASS. Several participants wrote about having very context-specific symptoms but having to compromise their reply to questions because the response options offered were too superficial or generalised. In relation to this, participants raised criticism of being unable to clarify their responses with more details. One individual stated they were “not able to communicate contributing (background) circumstances” and there was therefore a “strong likelihood of diagnosis beings skewed”. Another expressed disappointment that certain questions “did not allow for answers that needed a "depends on the situation" or a "sometimes" response”. Some participants reported being uncertain about what questions were asking, which for one individual

"lead to room for inaccurate answers and therefore results". Approximately 7% of participants referred to the e-PASS as being too long and time consuming. One individual related the length to an increased likelihood of dropping out. The least reported disadvantages of the e-PASS referred to technical issues such as poor internet connection, the lack of engagement with the program (e.g. due to distraction), and perceived inaccuracy.

5.4 Discussion.

There are few empirical studies of the user experience of web-based clinical assessment despite its potential importance in informing the utility of this assessment form. Hence, this study aimed to address this by examining the user experience of the e-PASS. Of particular focus was how people viewed their experiences of the e-PASS in terms of key areas relevant to the online and clinical assessment literature. To help with identifying the strengths and areas requiring improvement, participants were asked to identify liked and disliked aspects of the e-PASS. Of additional interest of this study was how the e-PASS compares with more traditional assessment methods, namely the CI conducted in Study One. The e-PASS is relatively unique in its appearance, content, structure, and context of use, which could influence how people perceive and experience the program. Hence, this study was conducted without any clear expectations, though results will be discussed in relation to relevant literature. As the e-PASS shares some characteristics with general web-based programs, the findings of this study may be generalisable, though should first be replicated to confirm validity.

5.4.1 Preference.

The study found equally high acceptability towards the e-PASS and CI based on peoples' willingness to do it again if needing another assessment. A lack of preference was also found across different groups considered in this study, such as those with and without a particular disorder or certain socio-demographic backgrounds. These results contrast with previous studies of computerised assessment, which have observed generally greater preference for a computer over interviewer administered psychiatrist assessment (e.g., Angle et al., 1978; Carr et al., 1981; Greist et al., 1981) or vice versa (e.g. Blouin et al., 1988; Erdman et al., 1992; Kobak et al., 1988; Peters et al.,1998). However, results of the present study are similar to those of Erdman et al. (1992) who found a similar preference between the computer and interviewer administered Diagnostic Interview Schedule (DIS) amongst a sample of psychiatric inpatient and outpatient clients. The only previous study which has examined the acceptability of a web-based program found that almost all participants would recommend the program to others (Parker et al., 2013).

While the present study did not address this, the results similarly indicate that actual program users are generally satisfied with their web-based assessment experience. However, there were various positive and negative aspects to the e-PASS experience, particularly when compared to the CI experienced in Study One.

5.4.2 Convenience.

As expected, the e-PASS was found to be convenient and easy to access, a quality which consumers particularly liked and viewed as an advantage of the program ("...able to be done at home in my own time"). Furthermore, the convenience of the e-PASS was rated with more agreement by those currently accessing mental health services compared with non-service users, suggesting a greater appreciation of the e-PASS's convenience when viewed alongside traditional service access. These results reflect the argued benefits of accessing online assessment over traditional assessment as well as offering online resources in general (Eysenbach & Kohler, 2002). They also contrast with Van Ameringen et al.'s (2010) study which found relatively low endorsement of the MACSCREEN (a similar program to the e-PASS) as a more convenient alternative to seeing a health professional. Interestingly, the e-PASS experience did not seem to differ considerably from the CI experience in terms of convenience. This could be due to the CI occurring over the phone and at a time and location of convenience for participants. A greater contrast in convenience may have been more noticeable if the CI had, for instance, been a face-to-face interview in a clinic setting.

5.4.3 Comfort.

The e-PASS experience was found to be only slightly more comfortable than the CI. This result extends on those of Erdman et al. (1992) and Kobak et al. (1994), who found no significant difference in reported comfort during a computerised and interviewer administered structured interviewer. However, the little difference in perceived comfort between the CI and e-PASS was likely because participants could undertake the CI at their preferred time and setting (with phone connection), which could explain the high endorsement of feeling comfortable during the CI, and to a similar extent as experienced during the e-PASS. Again, the e-PASS may have been viewed as significantly more comfortable if the CI was conducted face-to-face in a clinic setting.

While associated results should be interpreted with caution given the limited sample size, factor analysis found a negative relationship between rated comfort during the e-PASS and treatment motivating, indicating that less comfort (or more awkwardness) experienced during the

e-PASS correlated with decreased motivation for accessing a health professional and treatment after the e-PASS. Although it is possible that an adverse experience of the e-PASS could consequently deter follow-up service access, predisposing factors (e.g. underlying anxiety symptoms, personality, and treatment motivation) could also help explain the association. For example, a person with a more avoidant personality could feel greater discomfort or find the e-PASS more confronting, while having a stronger aversion of health professional contact, compared with a less avoidant person. Therefore, it remains unclear whether greater discomfort during (and perhaps elicited by) the e-PASS has a direct causal effect on subsequent consultation and treatment access.

5.4.4 Engagement.

The study also found users generally took the e-PASS seriously, were not distracted, and responded to items with care. This extends on the results of Erdman et al. (1992) which found inpatient and outpatient service users were generally compliant and engaged whilst performing a computerised version of the DIS (Erdman et al., 1992). The finding that the e-PASS was taken seriously is also promising given that the online experience can be associated with negative disinhibited behavior (Suler, 2004) and disengagement due to boredom (Cunningham & Van Mierlo, 2009). It also helps mitigate concerns that extraneous factors (e.g. distraction) and the lack of control over the testing environment could compromise the behaviour of web-based assessment programs (Barak & Buchanan, 2004). On the other hand, it is perhaps not surprising that participants were fairly engaged in the program given they chose to perform the e-PASS and participate in this research.

However, results also indicated moderately better engagement with the CI than the e-PASS in terms of taking it more seriously and answering with care, added concentration, and finding it more interesting and engaging. This is not surprising given that e-PASS interaction is generally less responsive and dynamic than when speaking with an interviewer, who can perhaps elicit a greater sense of attention and accountability than a computer. Also, as Bowling (2005) notes, responding to an interview involves less demanding cognitive tasks than a computerised questionnaire, which could partly explain why people found it easier to concentrate with the CI. As one might expect, individuals stating a need for assessment were found to take the e-PASS more seriously as well as more easily concentrate and be less distracted, suggesting that existing assessment motivation can facilitate engagement with web-based assessment. This is consistent with Ritterband and colleagues' (2009) model of behavioral change which posits that motivation,

as a user characteristic, has a distinct role in the outcomes of an online intervention program. More specifically, it reflects the notion that people will display a more positive response style when an online measure produces an outcome that helps meet their needs (e.g. personal development, treatment recommendations; Naglieri et al., 2004).

5.4.5 Ease of use.

The e-PASS was found to be generally easy to use, with high agreement in the statement ratings ("it was easy to use"), and about 10% of the sample referring to this as a particularly liked aspect of the program. Comments about this were very brief, though a few people referred to the "ease of entering data" by clicking responses. While these results support the overall usability of the e-PASS and suggest that the questionnaire format can be appropriate method of web-based assessment, the study also found some people experienced specific usability issues, particular regarding the response method, the length, and the impersonal nature of the program.

5.4.6 Limited response options.

The most apparent criticisms of the e-PASS's usability concerned its response method, which was seen as "restrictive" in not allowing people to express and clarify responses, select alternative response options, and modify previous responses. This aspect was frequently referred to as a disliked aspect ("no room to express opinions") and a perceived disadvantage of doing the e-PASS ("couldn't clarify my answer"). It was also viewed as distinct point of differentiation between the e-PASS and the CI, as participants considered the e-PASS more restrictive in expression and "narrow in exploring for symptoms" than the CI. Erdman and colleagues (1992) similarly noted respondents felt better able to describe their thoughts and feelings to an interviewer-administered DIS compared with a computerised version. These results are understandable given that interviews are generally more flexible in allowing participants to elaborate on responses as part of attempts to establish rapport (Groth-Marnat, 2009). In the present study, the CI asked many closed response questions (e.g., yes/no) but allowed people to discuss the context of their symptoms (e.g., as required in ADIS-IV open questions), while the e-PASS predominantly used closed response items.

Issues with closed response items have also been observed with traditional instruments. For example, Pinninti et al. (2003) surveyed respondents of the MINI interview schedule and found several participants had difficulty answering yes/no to some items. Using closed items is

recognised as having both strengths and weaknesses in gathering information (Schwarz & Oyserman, 2001). In the e-PASS, closed items were used to resemble interview schedules such as the MINI, and because they are efficient and amenable for computerised branching, scoring, and data collection. However, the question arises as to whether the limited response options could have influenced the quality of e-PASS results.

5.4.7 Length and comprehensiveness.

While general ratings indicated disagreement with the statement that the e-PASS was "too long", the lengthiness of the e-PASS was a commonly disliked ("very long for an online survey") aspect of the e-PASS experience and was associated with comments about the program being "time-consuming" and "boring". Furthermore, participants tended to agree more with the e-PASS as being "too long" as opposed to the CI, despite the CI involving similar items and generally taking longer to complete. Other studies have also observed computerised assessment programs as being lengthy in comparison to clinician administered interviews despite sharing the same content (Peters et al., 1998). One reason why participants of this study viewed the e-PASS as relatively long could be that users did not expect the e-PASS to be as long as it was, given that the majority of the sample had not completed a program similar to the e-PASS.

On the other hand, the comprehensiveness of the e-PASS was often referred to as a *liked* feature of the e-PASS, which suggests some people value a more in-depth program. For those who had completed previous traditional assessment, almost a quarter of the sample highlighted the e-PASS as more comprehensive and/or in-depth. This reflects suggestion in the literature that traditional practice of assessment in the community, such as in primary care, can be narrow in covering different potential disorders (Zimmerman & Mattia, 1999), while structured interview schedules of which the e-PASS is modelled upon are more thorough, yet not usually made available to consumers. However, given the issues of length above, there appears to be a need to balance the benefits of a more comprehensive program (e.g., perceived user value as well psychometric properties) with the burden placed on respondents.

5.4.8 Impersonal.

The e-PASS experience was also commonly disliked and seen as disadvantageous for being impersonal and mechanical ("it felt robotic"). As an automated questionnaire which aims to be neutral and simple to use, the e-PASS presents very little except items in basic text using fairly neutral terms. Hence, this finding was reasonable, and is consistent with the observation that the

standardisation of online testing inevitably creates an impersonal experience which may not be appropriate for everyone (Naglieri et al., 2004). However, given the prominence of this issue, it draws attentions to Coyne, Bartram, and Smith-Lee's (2003) suggestion of balancing the cold medium of computers and the internet with the personal nature of assessment. While it is tempting to make the e-PASS more personal (e.g., adding a more human presence, as suggested by one participant), this may need to consider the impact on other important aspects of the e-PASS experience (e.g., perceived anonymity, privacy, judgment) apparently influenced by the lack of human presence.

5.4.9 Privacy, anonymity, and confidentiality.

The ability to perform the e-PASS on one's own was commonly favoured and seen as advantageous in promoting privacy ("I could do it privately") and anonymity ("I liked being anonymous"). These results reflect commonly reported online experiences (Christopherson, 2007) and reasons why people go online for mental health resources (Berger et al., 2005), suggesting they also contribute to the appeal of web-based assessment programs such as the e-PASS. There were several experiential themes associated with being alone that were seen as beneficial for users. Some participants described a sense of autonomy ("time to consider answers, no pressure to complete quickly because someone was waiting") which has been associated with computerised assessment (Greist et al., 1987) and previously identified as an important function of privacy (Pederson, 1997).

Participants also described feeling more at ease and comfortable because they did not have to "meet anyone new, go anywhere new, and face scary things". While this indicated the e-PASS's convenience, it also highlighted other underlying needs for privacy, such as solitude and isolation which as Pedersen (1997) found, are associated with a need to withdraw from social interaction, and contemplate/plan for the future on one's own. On the one hand, this preference of performing the e-PASS on one's own could be seen as positive characteristic in the context of self-help behaviour. However, it could also reflect a degree of avoidance which is symptomatic of many mental disorders (Hayes, Wilson, & Strosahl, 1996) and could be counter-productive to longer term recovery, especially if it prevents more effective assessment with a clinician.

Somewhat surprisingly, the e-PASS was rated only moderately higher that the CI in terms of anonymity, despite the e-PASS having no human contact. However, the smaller than expected difference could be due to the e-PASS being undertaken as part of a research project (with participants recognising their results would be examined as part of a wider study) and

because the CI was conducted on the phone with a relative stranger, which involves less interpersonal communication (e.g., nonverbal) and likely provides greater perceived anonymity than a face-to-face interview with a more identifiable and familiar person (Ben-Ze'ev, 2003; Joinson, 2001). Furthermore, the results revealed equally high levels of perceived confidentiality towards the e-PASS and the CI. While this indicates consumers' confidence in the ability of the e-PASS to maintain confidentiality, similar ratings to the CI may again reflect participants' recognition that both activities would be subject to the same conditions of confidentiality underlying the overall research project.

5.4.10 Self-disclosure.

An often noted outcome of anonymity and privacy is self-disclosure. Unfortunately, statement ratings gave little evidence of whether the e-PASS promoted self-disclosure of sensitive topics, as items intended to address self-disclosure seemed to be responded to in relation to overall expression of symptoms and issues of respondents. However several people described the e-PASS as advantageous in promoting more honest responding ("I felt I could be very honest") and it was suggested that sensitive matters were easier to disclose to the e-PASS than a person ("I felt comfortable disclosing faults to a non-person", "I provided more honest answers than I may have opened up to a psychologist"). While this is limited evidence, these results correspond with literature indicating that online assessment methods can enhance self-disclosure of sensitive topics (Davis, 1999). However, it is uncertain as to what extent the e-PASS encouraged more honest disclosure and how this might impact on the evaluation of the criterion validity of the e-PASS as in Study One. Given the small difference in perceived anonymity between the e-PASS and CI, there likely was a limited effect of self-disclosure in explaining differences in diagnostic results between the e-PASS and CI.

5.4.11 Impact.

Results suggested the e-PASS encouraged follow-up consultation to some degree, and at least did not deter people from seeking additional assistance. There were also indirect indicators of e-PASS feedback having some impact on consumers. Firstly, a higher number of e-PASS diagnoses were associated with a greater willingness to see a health professional. Furthermore, receiving e-PASS feedback was commonly cited (e.g., "it helped pin point specifics about myself") as a "liked" aspect of the program, suggesting that consumers got something out of their results. However, further investigation is needed to explore the impact of the e-PASS in detail. In general, however, the results of the present study touch on those of Van Ameringen et

al.'s (2010) in which over half of the sample stated an intention to seek further assessment from a health professional after completing the MACSCREEN web-questionnaire. The present results also reflect the findings of Parker et al. (2013) in indicating that while some people do not act on web-based screening results, others gain from them.

Compared to the CI, however, the e-PASS was found to elicit less motivation to see a health professional. This was despite the e-PASS providing diagnostic feedback and the CI providing no direct feedback. This could be explained by a number of factors. For example, the CI may have modelled a positive experience of an actual assessment with a health professional or, as statement ratings suggested, was trusted as more accurate than the e-PASS. Regardless, it seems a clinician-administered interview has somewhat more motivational influence on consumers than the e-PASS, though this needs to be explored further.

The results also suggest that an existing need for assessment will enhance the impact of the e-PASS results in terms of enhancing motivation to seek follow-up consultation with a health professional. Furthermore, a need for assessment was also associated with: a greater willingness to redo the e-PASS and take the e-PASS "seriously"; feeling less distracted and finding it easier to concentrate. Together these results suggest that individuals seeking assessment are more accepting of and engaged with the e-PASS, and get more out of it compared with individuals not needing an assessment. This finding suggests that e-PASS access should either be limited to those who are seeking assessment as it will increase engagement with the program, or inform users that e-PASS efficacy is dependent on how motivated they are to receive an assessment.

5.4.12 Group differences in UE.

Significant differences in rated e-PASS experience between particular groups were also found. There were some genders differences, with results indicating men finding it easier than women to disclose information to the e-PASS. These results, which resemble gender differences observed by Kays et al. (2011), suggest that the web-based questionnaire medium is perhaps more conducive for men to disclose their symptoms than women, who perhaps prefer a more open and interpersonal medium of discussing their issues such as an interview. Given the various barriers men face in discussing mental health problems (Galdas, Cheater, & Marshall, 2005) these results suggest online programs such as the e-PASS may be a useful option for addressing men's assessment needs.

Experiential differences were also found between those with and without certain clinical disorders. Experiential factors appeared relevant to the symptoms of the disorder. For example,

as one would expect, individuals with social phobia indicated feeling more awkward during the e-PASS than those without social phobia, while people with GAD viewed the e-PASS as less accurate and effective in collecting relevant information than those without GAD. Furthermore, participants reporting lower quality of life appeared to experience relatively more restricted expression, more feelings of boredom, more difficulties concentrating, and lower self-esteem after the e-PASS. These results suggest UE statement ratings may have effectively acted as symptom measures. Nevertheless they indicate that the e-PASS experience can be influenced by the symptom profile of a respondent.

5.4.13 Association between UE and diagnostic accuracy.

The study also found certain factors of e-PASS experience associated with the e-PASS's diagnostic accuracy (measured by agreement with the CI), though mostly with small to moderate effect sizes. Diagnostic agreement of MDD appeared to have the strongest association with UE statements. Specifically, individuals with diagnostic disagreement for MDD rated feeling more distracted, finding the e-PASS lengthier, and feeling more put off from treatment after completing the e-PASS than individuals with diagnostic agreement for MDD. However, the specific relationship between experiential factors and diagnostic agreement was unclear. Attention impairment and negative thinking styles are typical symptoms of depression (APA, 2000) and have been associated with low acceptance of computerised assessment and negative computer attitudes (Weber et al., 2002). These results suggest individuals with depressive symptoms are more susceptible to disengagement from the e-PASS which reduces e-PASS accuracy for MDD (in the form of diagnostic disagreement with CI) and deters follow-up treatment.

On the other hand, there was a mixture of clinical and non-clinical cases within the diagnostic agreement and disagreement groups. The group with diagnostic agreement for MDD had a larger proportion of non-clinical cases of MDD than the group with diagnostic disagreement. Therefore, lower ratings of distraction within the diagnostic agreement group may have been an artifact of having lower depressive symptomotology within the group. Hence UE statement ratings may have indirectly acted as a symptom measure. Alternatively, proportionally more non-clinical cases may have meant briefer e-PASS administration and therefore reduced lengthiness and distractedness. While the relationship is not clear, experiential factors such as distractedness nevertheless demonstrate small to moderate association with the e-PASS's criterion validity, and this is perhaps mediated by psychopathology.

5.4.14 Experiential differences as a predictor of criterion validity.

Another area of interest was whether experiential differences between the e-PASS and CI could predict the criterion validity results of the e-PASS in Study One. This was found to be only relevant for the diagnostic results of two out of the four disorders investigated: MDD and insomnia. Specifically, diagnostic disagreement for MDD was more likely when a respondent reported the e-PASS as more "boring and tedious" and capturing less useful information than the CI. This result could be directly interpreted as evidence that lower engagement and reduced information conveyed to the e-PASS compared with the CI contributes to diagnostic disagreement for MDD. While this explanation has some intuitive value, it should be viewed with caution given the lack of causal inference underlying the statistical results. The study also found diagnostic disagreement for insomnia to be more likely when respondents reported the e-PASS as less engaging and interesting, more thorough in exploring for symptoms, less awkward, and more anonymous than the CI. The range of factors in this association makes it particularly difficult to explain how they each influence diagnostic agreement between the e-PASS and the CI. However, the results at least provide some evidence that specific experiential differences between the e-PASS and CI may contribute to diagnostic disagreement.

5.4.15 Limitations.

A notable limitation of this study was the reliance on a non-validated survey. There are several mentioned scales in the literature measuring UE related dimensions such as affect, aesthetics, engagement, and relevance (see Bargas-Avila & Hornbaek, 2011). Part of the reasoning for not including these in this research was that they have not been standardised with an online clinical population. Existing scales were also considered too time-intensive and narrow in exploring the UE areas of interest (e.g., anonymity, impact of follow-up behaviour) in this research. Instead, the development and use of the UE survey in this study was aimed at eliciting a more general representation of the UE of the e-PASS in comparison with the CI. However, the results, particularly of the rated statements, may not be entirely valid or comparable with normalised data given the lack of standardisation. Further research is needed to validate the UE survey and confirm the results of this study, while future UE research of online clinical diagnostic programs could consider using more standardised measures of UE.

Further to the issues of the UE survey, many participants seemed to rate similarly between corresponding statements referring to the e-PASS and the CI, as evident by the small differences in average ratings. While this was assumed to reflect actual similarities between the

e-PASS and CI experience, it may have been an artefact of the consecutive sequencing of e-PASS and CI related statements. This format was intended to encourage respondents to directly compare and contrast their e-PASS and CI experience. However, it may have instead confused participants and/or prompted them to respond similarly across the e-PASS and CI related statements. To mitigate this potential issue, the task of rating the statements could have been split into two, with the rating of e-PASS related statements performed immediately after the e-PASS, and the rating of the CI related statements after the CI.

Approaching participants immediately after the e-PASS to undertake the UE survey could have also reduced the risk of response issues associated with retrospective data (e.g. memory bias, false recall). The UE survey, was on average, completed a considerable amount of time after administering the e-PASS (mean average 13.8 days). The delay was partly due to participants having to firstly undergo the CI (a mean average of 10.4 days after the e-PASS) before being invited to do the UE survey. Requiring participants to experience both the e-PASS and CI was advantageous in allowing a comparison of participant's experience via the e-PASS and CI related UE statements. Again, the e-PASS related UE statements could have instead been presented to participants immediately after the e-PASS. This approach could have also increased and diversified the sample by enabling anyone who had done the e-PASS (rather than the subset of participants who had also completed the CI) to undertake the UE survey. A larger sample may have also produced more enlightening exploratory factor analysis results. While indicators supported the validity of resulting factors, the sample size in this study was arguably insufficient in meeting minimum requirements for robust principle component analysis.

Another issue with the structure of the UE survey was the ordering of the UE statement rating task before the open questions (e.g., liked/disliked aspects of the e-PASS experience). As the UE survey commenced with the UE statements, participants may have been primed to respond to open questions (e.g., regarding the liked/disliked aspects of the e-PASS) in ways that reflected the content of the statements. For example, reading and rating the statement "I felt anonymous" may have made themes of anonymity more salient in participants' minds, therefore increasing participants' likelihood of noting this as a liked or disliked, and advantageous or disadvantageous aspect of the e-PASS.

While there was some attempt to identify UE themes contributing to diagnostic disagreement for particular disorders, analyses did not simultaneously control for diagnostic disagreement on other disorders. As there was a high level of e-PASS and CI co-morbidity, examining the UE for diagnostic disagreement of one disorder may have been too specific, given

that an individual may have also experienced diagnostic disagreement for one or more other disorders. An alternative approach to investigating the relationship between UE and diagnostic agreement could have compared UE statement ratings with the number of diagnostic discrepancies for an individual. This could reveal a more significant link between e-PASS UE characteristics and higher levels of diagnostic discrepancy.

Regarding sampling limitations, the recruitment of actual e-PASS users was intended to obtain ecologically valid experiential data. However, compared with the larger (and presumably more representative) sample in Study One, this study's cohort had a higher proportion of participants with post secondary education. This demographic group may have been relatively drawn to this study because they were more willing to undertake the task of completing a survey. However, higher education may have influenced how people experienced the e-PASS (e.g., comprehending the items, navigating through the program) thus potentially biasing results.

Also, in comparison to the general population, there seemed to be a high rate of past and present treatment access amongst the sample. The impact of this on individuals' experiences of the e-PASS was noticeable, with various participants highlighting the contrasts between aspects of the e-PASS experience and their relationship with a health professional. Those with previous experience with mental health resources may have a better understanding of their symptoms and could more easily recognise e-PASS items. This was apparent in several participant descriptions of the e-PASS as being relevant and accurate in its line of questioning. On the other hand, those with previous mental health experience may have stronger preconceived views of their diagnosis, which may have influenced how the individual responded to items, as well as their experience of the e-PASS feedback. While the sample reflected current users of the e-PASS, further evaluation of the e-PASS with other population groups (e.g., non-clinical, non-internet users) is needed determine whether the findings of this study can be generalised across the wider population.

Also, it is possible that participation in this research influenced people's experience of the program and their reports of it. The study took efforts to minimise the influence of research participation, such as allowing people to complete the e-PASS in their chosen setting, and collecting data through an online survey as opposed to more direct approaches (e.g. observation). However, there is the possibility that recruiting participants prior to the e-PASS may have affected how they approached and responded to the program. For example, real-world users of the e-PASS could in general experience a greater sense of anonymity, privacy, and confidentiality because they know they will not be later contacted by an interviewer (as in Study One) or be asked to detail their experiences of the program. It would be interesting to see

whether e-PASS experience would have been perceived differently had participants been recruited directly after completing the e-PASS.

5.5 Conclusions

This study provides initial evidence of the e-PASS experience as perceived by actual users of the program. There was support for the overall acceptance and usability of the program, with many users highlighting the convenience, comprehensiveness of assessment, and anonymity as positive aspects of the e-PASS experience. However, there were also issues that detracted from the e-PASS experience, such as the perceived lengthiness and impersonal nature of the program, and the limited response options which constrained expression of one's issues. The e-PASS experience was also found to vary for particular groups of users. Those who were seeking an assessment were generally more engaged during administration and motivated to take further action after the e-PASS. Users with current mental disorders and lower quality of life were also found to have a more negative experience of the program. Experiential factors relevant to the diagnostic accuracy of the e-PASS were also investigated, with factors such as boredom and discomfort during the e-PASS displaying an association with the diagnostic discrepancy between the e-PASS and the CI. While these are preliminary results, they suggest that aspects of the e-PASS experience may contribute to the limited criterion validity of certain disorders observed in Study One.

Although revealing important facets of the e-PASS experience, there are some issues and questions that were not addressed in this research. For example, it remains unclear as to why people access the e-PASS and how they specifically perceive it before, during, and after completing the program. Also, there were suggestions that self-disclosure of personal information was enhanced during the e-PASS; however, the extent of this and its influence on the accuracy of the e-PASS was not clearly apparent. Last but not least, this study did not closely investigate the impact of the e-PASS program in areas such as emotions, beliefs (e.g., about one's issues, the value of web-based assessment), mental health literacy, and follow-up behaviour. Although the e-PASS may show signs of acceptability, further research is needed to understand its potential effect on users, especially if the program is to be used as a form of self-help intervention.

6.0 Chapter Six: A Qualitative Study of the e-PASS User Experience (Study Three)

6.1 Introduction.

Previous empirical reports of web-based clinical assessment programs have largely focused on psychometrics properties, with little attention to the user experience of the program. There is literature examining the usability and experience of computerised assessment, however, results have generally been quantitative in nature, focusing on specific issues (e.g., overall acceptability, comfort) and the frequency/prevalence of these issues. Study Two similarly investigated the user experience of the e-PASS using a largely quantitative approach to maximise the generalisation of the findings. The study pointed towards specific e-PASS characteristics and experiential themes of varying importance and relevance to actual e-PASS users. However, the data collection method (online survey) and its specific items resulted in a narrow representation of what is likely to be a complicated interaction of factors underlying the e-PASS experience.

Furthermore, there are aspects of the e-PASS experience which remain unclear, yet may be of significance in reflecting the value and function of the e-PASS, as well as web-based clinical assessment questionnaires in general. For example, an area that was not explored in Study Two was why people access the e-PASS and what perceptions they have of the e-PASS prior to performing it. Another issue that could be further clarified is how the e-PASS's diagnostic feedback and recommendations affect respondents in terms of their mental health understanding and their follow-up behaviour. Importantly, it is of interest as to what a program such as the e-PASS offers for people in the context of experiencing mental health issues.

To elaborate on these issues and provide a more in-depth investigation of the e-PASS experience, a qualitative approach appeared apt. According to Elliot, Fischer, and Rennie (1999), the general aim of qualitative research is to understand and depict the experiences of individuals as they approach, interact with, and live through situations. There are numerous benefits of qualitative research that have lead to its increased use in psychological research in recent decades (Krahn & Putnam, 2008). Unlike quantitative methods, qualitative research is less concerned with quantification, hypothesis testing, and predefined categories, and is more interested in exploring unique lived experiences by appreciating individual variations, meaning, context, and culture. In doing so, qualitative inquiry can provide a richer account of a phenomenon. Therefore, the aim of this study was to conduct a qualitative investigation of people's experiences of the e-PASS. It was intended that this would help elaborate some of the findings from Study Two, as well highlight more subtle yet potentially meaningful individual

experiences that were not elicited through the online UE survey as employed in Study Two. The central question underlying the qualitative investigation concerned how people approached, experienced, and responded to the e-PASS in the context of their circumstances.

6.2 Methodology

6.2.1 Qualitative approach.

There are several major approaches to conducting qualitative research, including grounded theory, ethnography, discourse analysis, phenomenology, and interpretative phenomenology analysis (Hsieh & Shannon, 2007; Krahn & Putnam, 2008). These share the overall aim of enriching the understanding of phenomena, yet differ to varying extent in philosophical origins, theory, and methodology (e.g., process of collecting and analysing data). For example, grounded theory is an intensive and systematic approach intended to develop new theory and research questions through critical analysis of all relevant data (Glaser & Strauss, 1967), while ethnography focuses on acquiring a deep understanding of social groups (e.g., the thoughts and behaviours of ethnic and cultural groups) often through extensive fieldwork (Krahn & Putnam, 2008).

On the other hand, phenomenology concerns the study of lived experiences by understanding how people perceive and make sense of an experience (Krahn & Putnam, 2008). It is considered a powerful approach to understanding the subjective experience, including an insight into individuals' thoughts, feelings, actions, and motivations (Lester, 1999). Unlike grounded theory, it generally begins with a research question and is concerned with describing experience rather than explaining it with deduced theory (Bowden & Green, 2012). As the present study aimed to reflect the experience of the e-PASS, a phenomenological approach appeared appropriate and was adopted.

Specifically, an interpretative phenomenological framework (IPA) to data collection was referred to. IPA is a relatively new qualitative approach with growing application, particularly in the health psychology discipline (Reid, Flowers, & Larkin, 2005). IPA can be considered phenomenological in that it focuses on how participants' make sense of their "lived experience" (Smith & Osborn, 2008). In line with general phenomenology, IPA typically explores research questions at an idiographic level and is an inductive approach which extracts meaning and themes from the collected data (i.e. "bottom-up") rather than test hypotheses or makes prior assumptions (Reid et al., 2005).

Inherent in the IPA framework, and a distinguishing element from a more general phenomenological approach, is the emphasis on the interviewer playing a role in the interpretative process (Michie, Smith, Senior, & Marteau, 2003). Extending on the Heideggarian philosophical underpinnings of phenomenology, IPA recognises that a direct account of someone's experiences is not possible, because it is influenced by the researcher's subjective interpretation which itself is influenced by the researcher's background (e.g. values, biases, knowledge, experience). Smith and Osborn (2008) refer to this process as a double hermeneutic, which describes how the researcher, through reflection, ultimately must makes sense of the participant's attempts of making sense of their world. On the other hand, unlike other forms of phenomenology (e.g. Giorgi, 1997), IPA does not advocate critical interpretation of what participants report; rather, implication and findings are largely based on what participants say (e.g. direct quotes; Pringle, Drummond, McLafferty, & Hendry, 2011).

It is worth noting that IPA is an open and adaptable framework (Smith & Osborn, 2008; Smith et al., 2009), which extends on the notion that phenomenological research methods tend not to prescribe fixed stages and procedures (Pringle et al., 2011). As a result, there are variations of IPA as well as phenomenological frameworks in general across the literature (Finlay, 2009). This study aimed to follow an IPA orientation; that is, it was guided by IPA principles in focusing on the subjective experience while taking into account the interviewers role in interpreting results. Methodological considerations, such as sampling, procedure, and analytical approach, drew upon suggestions in the IPA/phenomenology literature. Importantly, this study did not aim to produce findings that could be inferred onto the e-PASS user population. Rather, the results highlight e-PASS experiences of a small and select group of people, viewed through the subjectivity of this researcher who also performed the interviewing and majority of analyses.

6.2.2 Participants.

Considerations.

As generally recommended in phenomenological research, the sample was limited to a small number of participants (Smith & Osborn, 2008) with the aim of allowing some variation in observed experiences across participants; whilst producing a manageable amount of data to allow in-depth analyses and nuanced findings (Smith et al., 2009). Reid and colleagues (2005) reviewed the IPA literature and found that the mean sample size was 15; hence, this study aimed to recruit a similar number.

Recruitment.

It is important to note that phenomenological approaches reflect an idiographic form of inquiry which is focused on the experiences of individuals rather than representing a population. As such, recruitment did not aim for random or representative sampling (Smith & Osborn, 2008). While variation in the sample can result in more diverse reported phenomena, a degree of homogeneity in the sample is recommended to allow some shared experience (Smith & Osborn, 2008). Hence in this study, purposive sampling focused on individuals who had completed the e-PASS in a setting of their choice. Other commonalities amongst participants consisted of being adults, Australian residents, and willing to share their e-PASS experiences via a phone interview.

Sample information.

Potential participants consisted of individuals who had participated in Study One and completed both the e-PASS and CI. They were invited at the end of the CI to undertake this qualitative study as an alternative to the UE survey of Study Two. The period of recruitment took place between December 2010 and June 2011. Of the 30 people invited, 15 people (6 males and 9 females) agreed to participant in the study (Table 43). This sample had a mean age of 33 years (*SD* = 11 years), and received a mean average of 2.80 (*SD* = 2.95) number of e-PASS clinical diagnoses, and 2.20 (*SD* = 2.18) number of CI clinical diagnoses. Background information collected prior to the e-PASS indicated: over half of the sample (*n* = 9/15) endorsed a need for a psychological assessment; most participants resided in metropolitan (*n* = 12/15) or regional (*n* = 3/15) areas; most participants were employed full (*n* = 7/15) or part time (*n* = 6/15); all had completed secondary schooling; most were born in Australia (*n* = 12/15); and almost all had a history of accessing a formal service for a mental health concern (*n* = 14/15).

Table 43
*Summary of Details of User Experience Interview Participants**

Name	Gender	Age range	Seeking assess.	Main reason for e-PASS use*	Primary concern	Current health professional
Alex	M	35	Yes	Online treatment	Depression	None
Anna	F	27	No	General info.	Personality disorder	Psychiatrist
Bec	F	28	No	Online treatment	Eating disorder	Psychologist
Claire	F	45	Yes	Assessment, online treatment	Depression	Psychologist, social worker, GP
Dave	M	36	No	General info.	Anxiety	None
Deb	F	19	Yes	Assessment	Anxiety	GP
Heather	F	31	Yes	General info., assessment	Anxiety	GP
India	F	50	Yes	General info.	Anxiety	None
Jake	M	40	Yes	Assessment	Depression	Psychologist
Joe	M	29	No	Online treatment	Anxiety	None
Kristy	F	39	No	Online treatment	Anxiety	Psychologist
Leo	M	19	Yes	Online treatment	Anxiety	GP
Lisa	F	25	Yes	Assessment	Eating disorder	None
Mary	F	30	No	Online treatment	Eating disorder	GP
Richard	M	42	Yes	Assessment	Anxiety	None

Note. * Details of name and age have been modified; ** Reported during interview

6.2.3 Materials

The user experience (UE) interview.

A semi-structured interview schedule was created to guide the UE interview. A semi-structured format is generally recommended in phenomenological investigation as it facilitates rapport and the probing of interesting areas brought up by respondents, and encourages respondents to take the lead in discussing interests or concerns (Smith & Osborn, 2008). However, semi-structured interviewing also implies the interviewer has some understanding of the investigated phenomena and will apply some (though as little as possible) preconceptions during the interview (Krahn & Putnam, 2008). Scheduled interview items (see Table 44 for examples) were typically neutral, open-ended, and followed-up by unscheduled prompts from the

interviewer to clarify what the respondent meant (Krahn & Putnam, 2008; Smith & Osborn, 2008). While interviews sought an overall account of e-PASS experiences, there were also specific areas of questioning to address the interests of this research, such as: why people performed the e-PASS; how people found the "look and feel" of the program; what they got out of doing the e-PASS; and whether they had any suggestions of how it could be improved. The interview schedule was reviewed by colleagues and pilot tested on a peer researcher who had completed the e-PASS. General feedback indicated the schedule was easy to comprehend, relevant, and appropriately structured.

Table 44

Examples of UE Interview Questions

1. Why did you choose to do the e-PASS?
2. Before doing the e-PASS, did you have any expectations of what the e-PASS might involve and could offer? If so, please describe?
3. Describe your experience of the e-PASS regarding:
    a. The "look and feel"
    b. Accessibility and the environment in which you completed the e-PASS
    c. How comfortable, willing, and able you were in disclosing mental health symptoms through the e-PASS
4. How did your experience of the e-PASS compare to other forms of assessment you may have experienced, such as an interview with a health professional, or some type of questionnaire?
5. Was there anything that you liked or disliked about doing the e-PASS?
6. What do you think could be improved about the e-PASS and your experiences of it?
7. What was your response to the e-PASS feedback?
8. Any other comments about your e-PASS experience?

6.2.4 Procedure.

Interviewer training.

The UE interviews were performed by the author who previously conducted the majority of the CIs (i.e., from Study One) with participants of this study. The use of one interviewer is preferred in phenomenological research as it promotes consistency in terms of how questions are

prompted and contrasted (Green & Bowden, 2012). As part of preparation, the interviewer underwent a university-presented qualitative research workshop and read guidelines in qualitative interviewing (Krahn & Putnam, 2008; Smith & Osborn, 2008; Taylor, 2007; Turner, 2010).

The interviewer strove to adhere to qualitative interviewing principles including the need to (Taylor & Bogdan, 1998): establish rapport; be open and flexible; not impose pre-existing assumptions onto the interview; be prepared to probe beneath superficial comments; clarify meanings and interpretation; minimise overly technical language; and to communicate in a clear and simple manner that can be easily understood by the participant (e.g., avoid double-barrelled, leading, and multiple consecutive questions). In line with a collaborative approach to interviewing (Arksey & Knight, 1999), the interviewer also employed strategies such as sharing control of the interview process, encouraging the participant to have an active role in discussion, promoting trust, and self-disclosure (Taylor & Bogdan, 1998). The interviewer performed several practice interviews before commencing actual phone interviews and received peer-supervision during the period of interviewing.

User experience telephone interview.

At the end of their CI (Study One), participants who expressed interest in undertaking the UE phone interview were informed of the main conditions of this activity: their call would be recorded to assist data collection and that all information provided would be treated with confidentiality; reported data would be de-identified and involve pseudonyms to protect confidentiality; the interview would be semi-structured and last approximately 15-45 minutes; participants were entitled to end the call or withdraw from the study at any time; there were no right or wrong responses and they were encouraged to say as much or little as they preferred during the phone call. Participants who consented to this were arranged a convenient time to be telephoned by the interviewer. Phone calls were conducted via a VOIP computer program and recorded as MP3 audio files using associated software.

All interviews commenced with a reminder of the conditions of participation. The interviewer then asked questions in reference to the interview schedule. .The schedule was not strictly adhered to in order for discussions to remain flexible and in part driven by the participant (Avis, 2010; Krahn & Putnam, 2008). Participants were usually prompted to elaborate on and clarify their responses, particularly when they diverged from specific themes covered by the interview schedule. As a method of data triangulation to verify the meaning of what participants

were saying, the interviewer often reiterated key points made by participants and asked the participant for confirmation or clarification.

Phone interviews were conducted a mean average of 13.07 days (*SD* = 7.48, min. = 1, max. = 27 days) after e-PASS completion. The duration of phone interviews ranged from 16 to 35 minutes, with a mean average of 25.5 minutes. At the completion of the interview, participants were thanked for their involvement. No further contact with participants was made. After all interviews were completed, interviews were listened to using computer audio software and transcribed into electronic Word documents which, together with the audio files, were stored in a password encrypted laptop.

Interviewer reflection.

The interviewer (the author) naturally displayed some inconsistencies despite efforts to adhere to guidelines and maximise the quality of interviewing. For example, after listening back on recorded interviews and re-reading the interview transcripts, it was apparent there were several missed opportunities where certain vague, yet potentially meaningful comments by respondents could have been explored in greater detail with follow-up probes. Also, there were instances when the interviewer could have allowed more extended pauses for both interviewer and participants to reflect and elaborate on their responses. The interviewer recognised the importance of this prior to interviewing, but found they were, at times, trying to reduce awkward silences and appear competent by moving onto subsequent questions or summarising responses. While the latter can be viewed as helpful (e.g. for verifying responses, demonstrating that the interviewer was listening to the respondent), it may have disrupted a participant's stream of thinking. Occurrences of these and other issues were noted and reviewed. In light of these issues, it is worth highlighting that regardless of training and experience, an interviewer's performance will inevitably have its limitations and areas which could be improved.

Qualitative analysis.

Following a phenomenological approach, data analysis occurred after all interviews were undertaken and conducted primarily by the interviewer/primary researcher. Several activities were performed as part of the thematic analysis of the interview data (Hsieh & Shannon, 2005). Firstly, interviews were transcribed and collated into a single document. The researcher read through the transcripts several times to help make "sense" of the data (Turner, 2010). Common experiential themes represented by consistent phrases, expressions, or ideas (Kvale, 2007) were highlighted

and interpretative notes of these themes were made alongside relevant text (Smith & Osborn, 2008). The entire transcript of a case was read and noted before moving onto the next case. Thematic material identified in one case was used as a reference for subsequently analysed cases to provide consistency in the thematic analysis across participants. A cross-case technique of displaying themed data (i.e. presenting statements of different cases adjacent to one another) was then used to assist in comparing examples of the same theme (Miles & Huberman, 1994).

Credibility in qualitative research.

A risk of qualitative research is failing to understand the context of experience and the relevant thematic categories, thus resulting in findings that do not accurately represent the data (Hsieh & Shannon, 2005). This can be considered an issue of credibility which resembles the quantitative paradigm of validity and reliability (Hsieh & Shannon, 2005). To establish credibility of the findings, several activities were undertaken to alleviate biases. Firstly, when drawing out experiential themes, source and method triangulation were used. Source triangulation is a common technique referring to the use of multiple sources from a data set to corroborate experiential themes (Baxter & Eyles, 1997). This is demonstrated by presenting quotations from different participants in support of a particular idea.

To further reduce bias, the process of reviewing and interpreting data involved consultation with peers for possible alternative perspectives (Smith et al., 1999). For several interview data sets, a second researcher reviewed the extracted themes and interpretations, in order to confirm appropriateness. When differences emerged, a third person was involved to help reach consensus. While the process of data triangulation is not perfect, it can improve the reliability of how data is interpreted (Mathison, 1988). Prior to finalising the themes and associated data, the author conducted a “verification step” which comprehensively reviewed the data for discrepancies, overstatements, and errors (Smith et al., 1999). Finally, the results of the study were written in consultation with the supervisors of the author to ensure they were relevant and meaningful to the research question.

6.3 Results.

6.3.1 UE qualitative interviewing themes.

The following sections describe the major themes identified from the qualitative interviewing data, some of which were included to address more specific research questions (e.g., what is the impact of the e-PASS?) of this study. Themes are exemplified by quotes

expressed by participants of the qualitative interviewing, with some individuals receiving more representation than others. When presenting the themes, variations within the sample were compared and contrasted using brief examples of quotes from different cases. This approach, which is consistent with the goals of cross case-comparison, were intended to help indicate the generalisability of a theme by showing how it is expressed across a number of people (Khan & VanWynsberghe, 2008). While the context of certain cases are not apparent, more detailed accounts are also included to preserve the uniqueness of particular cases in demonstrating certain (more complicated) themes (Khan & VanWynsberghe, 2008). Names have been changed (as consistent with Table 43) and minimal potentially identifying information is detailed in order to protect participant confidentiality.

6.3.2 Reasons for accessing the e-PASS

People reported trying the e-PASS and online resources as an alternative to traditional services. The free accessibility in contrast to expensive traditional services was often referred to as motivation. For example, one individual who had previous therapy stated she was "not in a position to pay a psychologist $130 per week". Participants also described the e-PASS as "something new" to try, without having any clear expectations. Deb for instance was referred to the AO website by a television segment and had never previously tried online resources. Anna, on the other hand, had a long history of using traditional and online resources and accessed the e-PASS "out of curiosity to see if it works or if it was crap, like a lot of shit that's out there". According to Anna, an "accurate diagnosis is the first step" to managing mental health, hence she often checked her diagnosis by seeing if "the same things were coming up" across different assessment methods.

A few individuals were more interested in undertaking a treatment program on AO and completed the e-PASS as a pre-requisite activity. As Joe stated, "I already knew what was coming...I was just looking forward to getting it over with and getting into the self help section". In such cases, as seen across many participants, it was apparent individuals had a well-formed view of their main mental health issues from prior experiences with services or self-help. Hence, many described a motivation to do the e-PASS as a means of confirming their understanding.

There were several accounts of negative experiences with health professionals which suggested a tainted view of health professionals underlying an interest in online resources as a means of self-help. One individual had previously seen two psychologists who "didn't really understand anxiety", so he turned to online resources (e.g., forums). Others spoke about feeling

unsupported and invalidated in their past health professional relationships. Richard, who reported strong somatic anxiety symptoms, said his doctor was “discounting” of his problems (“looks at me and thinks...he’s just making this stuff up again”) which prompted Richard to “look for another doctor with more mental health training”, and also lead him to the e-PASS. The tone across the sample, however, was not unanimously negative towards health professionals. In many cases, individuals discussed past interactions in contrast to more recent positive relationships. For example, Anna spoke of having a strong connection with her last psychologist (who she last visited a few months earlier), but knew more about her “than anyone else”.

6.3.3 Complimenting traditional services.

Indeed, some described a good working relationship with their current health professional, and accessed the e-PASS to compliment this. Anna, who saw her psychiatrist on a monthly basis, stated: “sometimes people can get bogged down about what their psychiatrists say...but I like to get lots of different opinions; it’s a way to get another opinion”. While cases like this reflected personal initiative in accessing the e-PASS, others were prompted by their healthcare professional. There were three distinct cases: one individual (Claire) performed the e-PASS in a clinic setting and was encouraged and assisted by her care worker; another individual was recommended the e-PASS by their GP; the third person (Jake) was referred to the e-PASS by their clinical psychologist as “homework” during their early work together.

6.3.4 A preference for privacy.

The e-PASS was mostly completed at home, with many stating a preference of completing the e-PASS in a private or personal setting. Dave, who had a relatively brief e-PASS experience and received few disorders, said he particularly liked being able to do the e-PASS in “the privacy of your own home”. Dave maintained focus while completing the e-PASS, despite receiving a phone call mid-way through. Alex also found he could concentrate, and avoided “multi-tasking” with other activities. Deb, who had completed “similar” questionnaires with a psychologist, said she found the e-PASS “easier to do...because you can do it in the privacy of your own” and because “you don’t have the pressure of people around you making judgments”. Only two individuals performed the e-PASS outside of home. As previously mentioned, Claire undertook the e-PASS in a clinic setting and ran out of time as the centre closed, hence had to finish the e-PASS at home. The other individual, Anna, performed the e-PASS in her work place, a mental health related organisation. Anna said she still felt comfortable as others around her would assume she was performing the e-PASS for work reasons.

6.3.5 General usability.

In general, people described the e-PASS as “easy to do”, because it was “organised”, “straightforward...and... just involved reading and clicking”. While the general interface was viewed as “standard”, and similar to previously completed questionnaires, this was considered a positive attribute as “it was familiar to what...people use”. The sequence and structure of questions was described as “self-explanatory”, with a number of people recognising the branching of screening items to more in-depth items. While one person liked how this helped “focus on the things you said yes to”, another suggested people could easily manipulate their results because the sequencing was predictable. The e-PASS feedback report was distinguished for being descriptive and in-depth, while one individual praised the design of allowing the expansion of text boxes and having key issues highlighted.

Naturally, there were reported usability issues of varying significance. Minor issues included: the screen not fitting a person’s monitor which was received with some frustration; having to open drop boxes to select a “no” response for certain items; the visual appearance of the program looking “a bit boring”; having to scroll down certain pages to view items; and the wide spacing between item responses which required “effort...to move the mouse across the page”. Lisa highlighted the inability to “push back” which resulted in her “accidently pressing the wrong thing” but being unable to “go back” and change her response. Claire could not link to her desired “self-help program” from the e-PASS results page, which left her feeling “a little stupid” and thinking the e-PASS was pointless because “all (she) was doing was answering questions”.

6.3.6 Lengthiness.

The most prominent usability issue reflected the length of the program. Some people were surprised and expected the e-PASS to be shorter as experienced with other questionnaires. The e-PASS’s length was occasionally viewed in positive terms (e.g., comprehensive, in-depth), and preferable over less specific questionnaires (e.g., K-10) which were described as “vague”. However, several saw the e-PASS as “too long” and associated it with negative feelings (e.g., “frustration”), disengagement (e.g., “boredom”) from the program, and more careless response behaviour. For example, Mary, who took 33 minutes to complete the e-PASS, said she “was so exhausted” that she “didn’t think about what (she) was saying”. Similarly, Claire, who undertook the e-PASS in a formal setting, said she had to “rush through it towards the end” and felt she “didn’t answer some of the questions properly” as a result. Despite issues with the length, participants persevered in completing the e-PASS. Jake summed this up by stating:

"I was thinking 'what, more?'...and I thought 'f### this', but then I thought 'well there's a good reason why they're asking this and if I'm going to do this, I might as well do it properly'. So in some ways, I tried to modify my behaviour...I tried to stick to it until the very end"

Suggestions were made to address the length of the program. These included shortening the e-PASS, though some individuals recognised shorter questionnaires as being less accurate. It was also suggested that the e-PASS give "a person a better view" of their progress, encourage breaks (e.g., "every 15 minutes"), and allow people to save progress and return at a later point.

6.3.7 Miscomprehension.

Comprehension issues were also experienced by several people. Heather said she "couldn't quite make sense" of some of the items because they were "worded in a strange way". Lisa said she had to "re-answer some of the eating...and sleeping questions" because she wasn't sure about their meaning. Similarly, India found that "sometimes the question were too wordy…a bit long winded" which required her to re-read certain questions. Along with several others, she also found items repetitive, which added to the perceived lengthiness of the program:

"...from memory, some of the questions were very similar, but repeated at different times. I don't know if that was meant to gauge response times, or to see if you would answer differently a second time around. So I think that's part of why it took so long."

Jake also "felt like (he) was reading the same question" but recognised that items in fact assessed "different components of anxiety". This helped him appreciate that the e-PASS was not "trying to be painful" and, as previously mentioned, motivated his completion of the program. India, on the other hand, said there were instances when she wanted to "speak to someone" so she could "say to them that I didn't quite understand" and "clarify what the questions meant". Given these difficulties, India felt she "wouldn't rely on a questionnaire" though acknowledged it is "probably beneficial in triggering a bit of reflection".

6.3.8 "A bit of discovery...a bit of confirmation".

Indeed, a commonly reported aspect of undertaking the e-PASS was of reflection and raising awareness of personal issues, though to varying degrees. Kristy did not initially see herself as "particularly depressive, and yet after doing it (the e-PASS), it became more apparent there are very clear indications of it". The experience was also psycho-educational for Kristy in terms of redefining her understanding of depression. Other individuals expressed gaining an

appreciation of secondary issues they had overlooked before doing the e-PASS. Heather exemplified this by stating:

"*I think I got a bit of a confirmation of what I thought I knew about my issues and at the same time, it brought to light some other aspects that I haven't concentrated on in my life; in terms of alcohol, gambling, and eating. It made me a bit more aware of these things, which I'm not normally aware of."*

As a result, Heather said the e-PASS provided ***"***a bit of discovery and a bit of confirmation" of what she already knew. Similarly, Richard said the e-PASS addressed "a few things that were new to me, like...the avoidance things". However, Richard said after reaching a "point where you've read it all", the e-PASS acted as a reminder of "all the stuff you've seen before". Alex, also an "experienced consumer" said the e-PASS "confirmed everything I've already seen, especially with psychologists" but still raised "a few things...to think about".

6.3.9 Impact on follow-up behaviour.

There were few instances of the e-PASS results having a major impact on follow-up behaviour and outcomes. The most significant was Jake's e-PASS experience. Jake was referred to the e-PASS by his clinical psychologist after three sessions together. Jake commenced the program with positive expectations as it was recommended as a "good diagnostic tool" which could assist in his assessment. However, he had a mixed experience completing the e-PASS, finding it frustratingly lengthy and repetitive, and at times confronting. Jake said he felt disappointed when he received his feedback as he thought it was vague and simplistic given the effort he put into the program.

Furthermore, Jake's results presented, amongst other e-PASS results, a clinical diagnosis of PTSD which he initially found "surprising", "stupid", and indicative of the e-PASS being overly "sensitive". However, after personal reflection and consultation with his psychologist, PTSD was deemed relevant and became the focus of further assessment and treatment. Anxiety had been neglected in Jake's previous psychological work. However, Jake said that he had come to "realise that rather than spend all this time working on the depression" he should "also be working on the anxiety because the anxiety is actually worse" and may have triggered his depression. For Jake, there was a sense of breakthrough and relief in being able to link a cause to his mental health issues, and this motivated him to work on his issues. In retrospect, Jake said he viewed the e-PASS as being "very accurate" and effective in asking appropriate questions which raised insight into his issues.

6.3.10 "It wasn't really a wake-up call".

Otherwise, as observed across a number of participants, there was limited motivation to pursue follow-up activities in response to e-PASS results and recommendations. As Bec stated, she "did nothing" after completing the e-PASS. For Deb, the e-PASS feedback "wasn't really a wake-up call" as she had "been hearing about these things for a while". Lisa also said the e-PASS did not make "much of a difference" for several reasons. Firstly, she had doubts as to "whether it was correct". Hence, she was "weary" about the significance of her results, and was concerned it would "put things in my head" and turn "possible problems" into a "real problem". Along with a few others, Lisa highlighted the "impersonal" nature of computer assessment as a reason why her e-PASS experience did not have a major bearing on follow-up behaviour.

6.3.11 Impersonal.

There were various references to the e-PASS experience as being impersonal. Anna "felt it could be a bit more comforting" as well as normalise the experience of mental disorders to encourage people to seek follow-up. Richard said he felt the e-PASS lacked "a human factor to make you feel there's help" after presenting its results, and that a supportive comment such as "congratulations, you made a big step towards your recovery" could encourage people to seek further assistance. For Richard, the e-PASS particularly lacked reassurance, which for him, was strongly associated with his anxiety:

"...because of my panic disorder or whatever, when you're going through these things...you're constantly looking for reassurance...that you're not going nuts or that symptoms you're feeling are not going to kill you"

For Richard, his need for reassurance was partly met by his doctor and better addressed by other websites with a more "human touch" (e.g. video of a consumer). Richard emphasised this when referring to another website featuring a central character that elicited his "loyalty" and willingness to "believe in him". India also expressed a need for follow-up support and reassurance to address the abrupt ending of the e-PASS and the process of reflecting on "memories":

"...I felt when you finish all that stuff you've been thinking through and bringing back all those memories; you finish the e-PASS and it just then cuts off. Then you think 'I got that result, and that's it'. It would be good to conclude somehow with some reassurance, or a reminder that a lot of time, it's ok…."

6.3.12 Need for support.

A need for support during and after the e-PASS was expressed in various terms, though largely in relation to the inadequacy of the program. Claire said her support worker (who sat alongside her for half of the session) helped clarify some items, which indicated her difficulty understanding the e-PASS on her own. Another participant stated she would have liked someone to contact her to explain the results, give advice, and refer to where she could go next for help. There were also references to needing emotional support. Jake reported feeling suicidal a few days prior to completing the e-PASS and noticed "a drop in mood" after completing the e-PASS. While he did not explicitly state the reasons, it was suggested this "drop" was to do with answering "confronting" questions about areas such as trauma and suicidality. However, Jake said he was fortunate to be able to "debrief" with his doctor later in the day, implying that his doctor provided a level of support which the e-PASS lacked.

Anna was concerned at how easy it was to report suicidal ideation and continue on with the program. In her case, she said she wasn't suicidal, but she "wondered" as to whether her report would be "followed up" which suggested an uncertainty about the supervision provided by the e-PASS. Anna was critical of the e-PASS's recommendation of a crisis phone service (i.e., Lifeline) as a means of crisis support. She was concerned this was not adequate because "the people on the other line aren't very well informed...and...might refer you to a service with a 6 month waiting list". Recognising a need for more adequate support in addressing suicidality, Anna suggested the e-PASS feedback provide detailed instructions and encouragement to accessing services such as a psychiatrist or therapist. Along these lines, Jake suggested different levels of recommendations based on how severe the risk of harm was reported.

6.3.13 Trust and credibility.

There were several mentions of trust and credibility towards the e-PASS. Richard, for instance described a sense of understanding from the e-PASS items, stating that "you feel like you're being asked questions by someone who knew what it was all about", which differed from his previous consultations with inexperienced health professionals. Mary said she "was trusting of it all" because "it felt pretty secure and safe" and "because it's from a university" so it's "not going to post rubbish". Alex was more specific in linking the credibility of the e-PASS to the Australian Government and unversity logo, which gave him "confidence" that the site was intended to "help...and gather research for the problems I have".

In relation to trust, there were some concerns about information privacy. Richard usually disliked being asked “who you are” because “you worry about who the organisation is and whether you want them to have your information”. However, he was not concerned about this with the e-PASS as “you just needed to make a username and password, and bang, there you go”. When asked about his response to the background questions, Richard said he did not mind doing them as they are “pretty standard these days”; however, he recognised that “people who are less trusting might”. Alex, on the other hand, questioned his privacy on a few occasions. In particular, he expressed concern about supplying email as a username as it could include identifying information. He also wondered whether having to provide “doctors details...might put some people off” though also acknowledged it could “ease some people’s concerns”.

6.3.14 Perceived accuracy.

The e-PASS was often considered more “in-depth” than previous forms of assessment and other online resources. Alex, prior to undertaking the e-PASS, was cynical about its accuracy and thought it would be like “...other online things which ask you a very limited amount of questions and give you some broad sweeping diagnosis that may or may not be right”. But he “was surprised about the level of detail” and was generally accepting of his e-PASS result. Another participant described previously completed questionnaires as “useless” for being too brief and generalised to provide “any meaningful outcome”; whereas the e-PASS was “much more specific as a diagnostic tool”.

Richard felt the e-PASS was more accurate than “trawling through the forums” and getting “advice from every random person on earth”. However, he was also “sceptical” about his “subclinical” e-PASS result because he could not “imagine feeling any worse” than he currently did. In addition to considering the e-PASS as “wrong”, Richard felt a sense of invalidation as his subjective experience of his symptoms was “brushed off”. As a result, Richard appeared ambivalent about his e-PASS recommendations.

Leo found the e-PASS was “quite in tune” with his thoughts, and “described a lot of feelings...quite accurately”. Leo was “surprised” about this because he expected, based on past experiences of questionnaires, to be “catering” to questions which “might miss something” as a consequence. Leo said his final results were “pretty expected”, though they “picked up things that weren’t really appropriate” such as his cannabis use, which Leo believed was not problematic. While Leo hinted that the e-PASS was overly sensitive (“it came up with every disorder

possible...") he appeared open minded about the results and said he could see how issues such as "body dysmorphic...relates to (him)".

Others, however, were more critical about the accuracy and usefulness of the e-PASS. Some simply described the e-PASS as "inaccurate" while others implied this by disagreeing with the diagnoses they received. Claire, for example, said she received a social phobia e-PASS result even though she is "pretty good talking with people". Claire stated that her main problems were instead MDD and GAD.

Anna was not surprised by the inaccuracy of the e-PASS as she had similar difficulties with other instruments. She referred to a previous measure which "threw up the same things" (i.e., as the e-PASS), many of which were ruled out by her and her psychiatrist in the past. Anna recognised how her presentation (which was previously diagnosed as borderline personality disorder) resembled PTSD and bipolar disorder but she understood it was difficult to "clarify" her symptoms and "tell a computer that". For Anna, the e-PASS made similar mistakes to other instruments and could not match the nuanced understanding of her issues as well as herself or her psychiatrist.

6.3.15 "An inaccurate interpretation".

False assumptions made by the e-PASS were often related to the inaccuracy of the e-PASS. A few people noted that the e-PASS over-emphasised issues that they believed weren't relevant. Substance related issues, in particular were raised as areas which people felt were being mis-interpreted as problematic by the e-PASS. In relation to this, Kristy said the e-PASS "makes a lot of assumptions that...felt like an inaccurate interpretation". Mary said she was "a bit annoyed" when the e-PASS presented items about addiction when she reported valium use taken while in hospital. Bec also found the e-PASS asked:

"...heaps of questions about marijuana use...making a much bigger deal that it actually is. For me, I use marijuana occasionally and socially, but it (the e-PASS) took it to be quite a thing".

Joe viewed himself as being over-diagnosed by the e-PASS, and attributed it to the over-sensitivity of e-PASS items. He felt this differed from the diagnostic process of a psychologist, who would start off with "vague questions...then ask more specific questions" if relevant. Joe referred to the e-PASS's gambling module as an example of how the e-PASS would present unnecessary items. Having endorsed the item about recently gambling, the e-PASS "caught onto" this and "highlighted all of the negative aspects" which he responded "no" to. Joe said these

subsequent items were unnecessary because he knew there wasn't a problem with his gambling. According to Lisa, this issue stemmed from the subjectivity of item interpretation, where "two things...never really mean the same thing". These issues of misinterpretation seemed to limit the view of the e-PASS's accuracy. For example, it was commented that e-PASS could not "adequately reflect the degree of which....things are problematic".

6.3.16 Expression.

Some individuals particularly disliked not being able to express the details and context of their symptoms and also associated this limitation to the inaccuracy of the program. A few noted not having sufficient response options, though did appreciate the option of typing comments in the "other field". Lisa thought some of the questions "were very rigid" which left little "room to move". Hence, for some items, she ended up answering a "specific way" rather than "give an answer (she) felt". Lisa went onto suggest this item specificity could result in the e-PASS overlooking problems. Anna was used to clarifying her responses (i.e. when being interviewed by someone) and therefore found the e-PASS relatively "frustrating". Like others, Anna felt she was compromising her responses by "choosing something you don't necessarily agree with, but it's the closest to what you've got". However, Anna acknowledged this was a limitation with computerised questionnaires. Kristy generally agreed with her feedback but wanted to "elaborate" on factors she believed contributed to her anxiety symptoms, such as her medication use and stressors. On the other hand, she could "appreciate that it would be difficult to fit it into the assessment, because you then need somebody there to read all of this, making judgment calls whether or not that is contributing to the anxiety problem".

For some, however, the constrained expression was not an issue. A few individuals preferred "pointing and clicking" and liked not having to "write out answers" which was a disliked aspect of previous questionnaires and seen as "laborious". Richard found the e-PASS mode of responding was sufficient, and didn't feel the need to say more "...because it wasn't really a therapy session. I was basically there to get an assessment of where I'm at." Similarly, Jake viewed the e-PASS as "a diagnostic tool" which "keeps to an objectivity that a discussion can't have" by being "only interested in circumstance". Referring to his disclosure of being sexual assaulted, he believed additional details (e.g. when and who perpetrated the act) were not needed in a diagnostic tool and instead more appropriate within a "proper clinical interview". In addition, Jake also felt the e-PASS asked "appropriate" questions, including questions "that other

people have never asked", but for Jake, were pertinent in helping him make a "connection" between his anxiety and depression symptoms.

Lisa questioned the self-report nature of the program and the predictable sequencing of items by commenting "it was easy for someone to manipulate, but I guess you're going to get that with any response". On the other hand, Deb pointed out it was in one's best interest to be as genuine as possible:

"There's no point doing these things if you're not going to be completely honest about your experiences, because the whole point of the program is to give you a better insight into what possibly is a course of treatment or actions you should take in relation to how you feel and your emotions"

6.3.17 Less confronting.

Undertaking the e-PASS was seen as less confronting than seeing a health professional, in part due to its limited questions. Anna, for instance, contrasted the safety of independently completing the e-PASS with the "terrifying" experience of seeing a psychologist for the first time, and being asked to disclose "everything about your life". For Anna, the e-PASS was viewed as less "invasive", because it did not involve "a person looking back at you saying 'how did you feel about that?'...".

For Jake, there was a sense of feeling confronted by certain e-PASS questions, but a greater comfort and willingness to address particular issues on a computer than with an interviewer. Jake described feeling "more in control" about what he disclosed to a computer. With a similar sentiment to Anna, Jake said "you just don't know what is going to be the next question from an interviewer", a statement which reflected his apprehension of the unpredictable nature of an interview, and having to discuss difficult topics (i.e., abuse) which would "put (him) through too much stress" and create a risk of "coming out...unravelled". Hence, Jake appreciated that the e-PASS focused on key information for diagnosis, sufficient to highlight "there is an issue...causing enough psychological distress", but without going into detail. This allowed him to disclose that he had experienced sexual assault to the e-PASS, thus leading to PTSD result, whereas he had withheld this during his Clinical Interview in Study One as well as with his clinical psychologist and previous therapists.

6.3.18 Less judgmental.

While the e-PASS was criticised for being impersonal, it was praised for not being judgmental. Jake noted that the e-PASS was different to talking to a human because it was a "faceless machine" which he found "a lot more reassuring" because he knew he was not going to be "judged". Deb, who did not regularly contact any health professionals, also described feeling less judged with the e-PASS, because she did not "have the pressure of people around...making judgments". As a result, Deb felt "more comfortable...than speaking to someone and answering their questions". One area Deb felt more at ease disclosing referred to suicidal ideation:

"...it's a personal experience, and there's a sense of embarrassment about it; so on the computer you don't worry about that embarrassment and you don't worry about whether this will change the person's point of view about me when I answer this question truthfully".

It was unspoken as to why Deb felt embarrassed about discussing suicidal thoughts. However, she referred to how, when first dealing with her mental health issues, it was difficult to be "honest...with how I was feeling" and that she "didn't want to just disclose it to anyone because I didn't want to feel embarrassed". For Deb, there was a concern that others "wouldn't know what it feels like", and therefore wouldn't "be able to help". On a computer, Deb did not have to worry about these issues. As "computers don't make judgments", Bec also found it was "easier to disclose private information to a computer" particularly regarding her binge eating symptoms (e.g., eating and compensatory behaviour). Bec also recognised that she would still be indirectly judged when her e-PASS results were reviewed by a researcher; however, because the e-PASS was performed "through a computer, you don't really make as much of a connection".

6.3.19 Stigma.

Despite a reduced sense of being judged, there were some concerns about the e-PASS results eliciting a sense of stigma. Dave felt "apprehensive" leading up to receiving his e-PASS outcomes as "most people don't want to be diagnosed as a nutter". These comments reflected a sense of stigma associated with diagnostic labelling underlying expectations of a program such as the e-PASS. Dave said he was "pretty relieved" upon receiving his results as he was "worried it was going to say something more severe". Leo received a report that listed "every disorder possible" with a mixture of clinical and subclinical results. He said this was a "good" feature because it "showed you're at risk of other things". However, he acknowledged feeling "a little bit" daunted by the number of diagnoses, and preferred the report to have "run through them one by one, rather than have this big list of things that are wrong with me". These comments suggested

that e-PASS feedback could be overwhelming and elicit a sense of defectiveness, especially when listing multiple diagnoses.

There was also concern that individuals less experienced with mental health services in particular might negatively perceive their results. For Anna, her e-PASS report, which listed "lots of disorders", would have been "quite worrying if (she) had never had any previous mental health professional experience" as "seeing all these things wrong with you" can be "scary" and overwhelming. Highlighting the potential harm the e-PASS could have caused if Anna had received it "five years ago" before accessing services: "...it would've stopped me from accessing services, because it would've scared me...and ingrained in me there is something wrong...or I would've obsessed over it". Anna could imagine what it was like to receive an "overwhelming diagnosis" as she was once diagnosed with borderline personality disorder by her psychologist and "sent on (her) merry way"; however, as she "had no idea what it was" at the time, the diagnosis overwhelmed her and prevented her from returning to therapy.

6.3.20 Usefulness and preferences.

The overall usefulness of the e-PASS was both supported and discounted. Alex viewed the e-PASS as a "great starting point" to "find some initial information" about one's issues and what might be of benefit. While he felt he gained little from the e-PASS and had adequate external support, he described the e-PASS as "something you could use to gauge your progress" and that he would return to the e-PASS in future. Dave praised the convenience of the e-PASS in highlighting how you can do the e-PASS "when you want, wherever you want" without having to "do it in a waiting room and pay a $180 fee at the end of it".

Leo said he would not perform the e-PASS again because "it's time consuming" but would consider retesting in a few weeks time "for some evidence" of the efficacy of his newly prescribed medication. Leo also stated a preference for the e-PASS over the phone interview because it felt "a bit more private" and "not judgmental" in contrast to "feeling a bit weird...talking about some of the weirder thoughts over the phone". Others, however, expressed reluctance to return to the e-PASS. Joe found the questions "too predictable" and "didn't see any point" to them given he could "identify (his) own symptoms, and match them with the symptom list" with the same accuracy as "doing the test". Joe recognised that "a lot of the questions were asking descriptions of the disorder" which "seemed like a 'no brainer'...because I already read the descriptions" on the AO website.

Given her adverse experiences, Claire recalled thinking "there was no way (she) would do it again" when asked via email to repeat the e-PASS for reliability testing. For India, an older individual with no major current mental health issues, the e-PASS was completed out of "curiosity" and without an expectation that the results "would lead anywhere" or provide "new information". Her issues during the e-PASS (e.g., item comprehension and lengthiness) confirmed her "thought that a questionnaire online, regardless of the result, is only one indication or one tool. And if you are really having problems with that particular topic, you would go and see someone about it". As such, she "wouldn't rely on a questionnaire" though acknowledged it is "probably beneficial in triggering a bit of reflection".

A preference for human assessment was associated with the need for interpersonal contact, the opportunity to discuss symptoms, and follow-up support which several individuals found lacking in their e-PASS experience. Certain individuals particularly valued the responsiveness of working with a human, as you are able to "get feedback straight away…if you're talking about something, you hear back immediately". For Joe, speaking to a health professional was also seen as a more effective way of accurately communicating one's issues, though at a cost of anonymity, affordability, and convenience.

A preference for face-to-face assessment was also associated with being more pro-active and committed about managing mental health issues. One individual reflected the value of human interaction in treatment motivation by stating that "human contact can't be ignored like the computer can" and "is what is needed to spur you into action". Claire, who preferred human interaction, said she liked making appointments because "you're more likely to follow through". Others also valued having "a specific appointment time with someone…and…a structure of things to do" because of how busy their lives were (i.e., "life gets in the way"). However, these individuals also described having strong current relationships with health professionals, which provided a basis for comparison with their e-PASS experience.

Anna also saw a limited need for the e-PASS because she already had "…things in place, like therapy twice a week and regular appointments with a GP". Anna described seeing her therapist as a means of maintaining "social contact" which continues to be important for her. Anna also valued her health workers capacity to collaborate with one another ("my psychiatrist would call my doctor to let them know if I was acting unusual") and provide more effective follow-up care. Lisa, who found the e-PASS impersonal and rigid, said she preferred a human interview over the e-PASS given her experiences of the CI in Study One. Lisa acknowledged initially feeling "a little embarrassed" at the beginning of her CI. However, "because of how the interviewer

responded", she found the interviewer "sweet and very personable" and gradually felt more at ease. Lisa also recognised the similarity in content between the e-PASS and CI but felt the questions were more "necessary" when presented by an interviewer. Overall, Lisa found the interview a "better experience".

6.4 Discussion.

This study aimed to address the lack of empirical research in the experience of online assessment (Barak & Hen, 2008) by interviewing e-PASS consumers' about their experience during and following the program, as well as their background and expectations prior to undertaking the program. To the knowledge of the author, this is the first entirely qualitative study of consumers' experiences of an online clinical assessment tool in a naturalistic setting. While there were some expectations based on previous literature related to computerised and online assessment/interventions, this study was also exploratory in considering the major issues relevant to people's experiences of accessing the e-PASS. The findings highlight differences between the experience of the e-PASS and traditional assessment, as well as variation in e-PASS experience across individuals and contexts.

In general, however, the e-PASS was found to appeal to people with prior experience of mental health services who accessed the e-PASS largely as a means of "checking" their mental health issues. This resulted in increased symptom awareness for some individuals but not others. The process of undertaking the e-PASS was generally easy to do, but was considered by some as lengthy, repetitive, and frustrating. While the e-PASS was seen as less judgmental and confronting in some respects, it was also viewed as presumptive and constraining in expression of one's issues. There were some reservations about the accuracy and usefulness of the e-PASS which contributed to the lack of follow-up behaviour. Overall, the e-PASS experience appears to consist of both positive and negative qualities which resemble the mixed properties observed amongst computerised assessment (Erdman et al., 1992; Peters et al., 1998).

6.4.1 Trust and perceived accuracy.

Consistent with the results of Study Two, the e-PASS program was associated with a sense of credibility (e.g., "written by people who understand anxiety"). One aspect underlying the e-PASS's trustworthiness was its link with an Australian university, which is consistent with findings that a website's credibility can be enhanced when it has an official source (Eysenbach & Kohler, 2002) and appears to be created by professionals (Klein et al., 2010). A reasonable level

of trust towards the e-PASS is expected from participants in this study given they voluntarily chose to access and complete the program. Trust may have also been enhanced by involvement of the other research activities, such as the CI and email contact from a registered Swinburne email account. Therefore, it is possible that actual e-PASS users may have a differing degree of trust when left to perform the e-PASS autonomously.

6.4.2 Perceived accuracy.

Despite expressions of general trust towards the program, there was skepticism towards the accuracy of e-PASS results. This reflects previous research indicating many internet users approach online material with caution as they recognise its limited reliability (e.g., Powell & Clarke, 2006; Dart, 2008; Lam-Po-Tang & McKay, 2010). Whereas Study Two results indicated people were generally trusting of the e-PASS's accuracy, this study demonstrated reservations amongst certain individuals stemming from aforementioned issues such as difficulties with comprehending and responding to e-PASS items and disengagement leading to careless responding. Given the potential influence of some of these factors on the accuracy of the e-PASS, and the limited accuracy of the e-PASS as seen in Study One, the caution participants have towards their e-PASS results is perhaps warranted.

As Lam-Po-Tang and McKay (2010) suggest, people's trust in online information is likely a complex phenomena. In their study of psychiatric patients internet use, less than one percent of the sample rated internet-derived information as unreliable, whilst their qualitative data suggested many more participants were aware of the risks. Similarly, the results of this present study highlight the many issues and concerns people have about their e-PASS performance which underlies an apprehension towards the accuracy of the e-PASS and perhaps web-based screening questionnaires in general.

Not surprisingly, a health professional was regarded by some as the best source for an accurate assessment. This result echoes the findings of Van Amerigan et al. (2010) indicating a lack of trust towards web-based questionnaires, and greater trust in health care professionals amongst consumers of a web-based mental health screening tool. It is possible that e-PASS users are influenced by the disclaimer at the beginning of the e-PASS which states that the e-PASS does not attempt to replace a full clinical assessment by a mental health professional. However, there was no reference to this by participants in this study as an influencing factor. On the other hand, it was apparent that seeing a health professional was not considered practical or necessarily beneficial by all, given the high costs involved and previous experiences of dealing

with health professionals who were unsupportive and lacking in mental health experience. These issues reflect common barriers to adequate mental health assessment and treatment (Prins et al., 2008) and partly explain why some people were willing to try online resources such as the e-PASS as an alternative.

6.4.3 Self-reflection

However, the study did find the e-PASS generally had a more subtle effect of encouraging self-reflection and raising people's awareness of mental health issues considered secondary to perceived primary problems. This is consistent with Barak and Buchanan's (2004) assertion that online assessment has a role in facilitating self-exploration and self-awareness particularly when applied as a self-help tool. In addition to providing diagnostic feedback, another aspect of the e-PASS which seemed to facilitate this was the range of items which helped respondents notice issues not previously recognised. A similar observation was made by Pinninti and colleagues (2003) who found members of their sample experienced greater recollection of their symptoms as prompted by items of the MINI interview. Also, in a study examining the experience of participants in psychiatric research (Marshall et al., 2001), the experience of responding to the SDI and other self-report questionnaire items were indicated as slightly to moderately helpful in promoting new insights.

6.4.4 Impact of the e-PASS results

A particular interest of this study was whether the e-PASS would impact on consumers' follow-up behaviour. As qualitative interviewing took place an average of roughly two weeks after the e-PASS, there was a sufficient period of time to examine at least the short term effects of the e-PASS. Generally, participants appeared largely unaffected and unmotivated by their e-PASS results with only a few mentioning they would or had already discussed their results with a health professional, and accessed a treatment program recommended by the e-PASS. The low rate of follow-up consultation in this study is consistent with reports that have shown few people discuss their online research with a health professional (Dart, 2008; Lam-Po-Tang & McKay, 2010). On the other hand, the lack of follow-up action appears to differ to a recent study by Parker et al. (2013) which found a large proportion of their sample taking some form of action (e.g., lifestyle change, seeking additional information, consulting with a doctor) after receiving a positive result from a web-based bipolar screening program. There were, however, noticeable limitations and differences between Parker et al.'s study and the current study.

For instance, Parker et al.'s (2013) study did not investigate whether these actions were prompted by web-screening feedback or were undertaken due to other causal factors, such as pre-existing treatment motivation or the influence of other resources (e.g., psychologist, another website) accessed during the same period. Also, the duration between screening and follow-up (3 months) in Parker et al.'s study was considerably longer than in the present study, which may have allowed more opportunity for people to undertake action. Finally, it is possible that participants in Parker et al.'s study were relatively motivated to undertake follow-up activities because the study targeted people specifically seeking bipolar testing, and because participants were informed about the 3 month follow-up survey immediately after completing the screening.

There was, however, one clear example in the present study (i.e., "Jake" who received a PTSD diagnosis) of the e-PASS experience resulting in several positive follow-up effects associated with diagnosis. In particular, the e-PASS results for this individual helped "name the problem" by providing a more beneficial framework of understanding their mental health issues (Pitt et al., 2009). It also legitimised personal aspects (i.e. trauma) and connected them to current issues in a way that was previously overlooked by health professionals. In response, it was evident that the individual experienced a degree of excitement, relief, as well as hope that the new diagnosis could lead to positive outcomes. This reflects findings from Haynes (2003) phenomenological study which found that a diagnosis can help make "visible the invisible" by reinstating a sense of life and the prospect of healing. Importantly, the e-PASS experience prompted this individual to consult with their psychologist, an uncommon outcome in this study, but nevertheless consistent with other studies that have shown computerised and online assessment can encourage people to discuss their results with their healthcare professionals (Chinman et al., 2007; Parker et al., 2013). While it is uncertain whether the e-PASS diagnosis will ultimately be efficacious for this individual, the results suggest it may provide a means to more appropriate treatment, which is another important outcome of a diagnosis (Pitt et al., 2009).

However, it is worth noting that the individual experienced various difficulties in the process such as a drop in mood immediately following the e-PASS, and initial reluctance to accept their e-PASS results. Ultimately, a confluence of factors likely facilitated the positive outcomes for this individual, including: their psychologist's recommendation of the e-PASS; the broad scope of the e-PASS; the online environment facilitating self-disclosure of trauma; and the follow-up support and further assessment provided by the individual's psychologist (and GP to some extent). Given the major role of the clinician, there is support for the observation in Lam-Po-

Tang and McKay's (2010) study that online resources could be more influential in health-related decision making when employed in a therapeutic context.

The limited impact of the e-PASS is perhaps not surprising given that many in this study's sample were sceptical of their results, not specifically in need of an assessment, and were completing the e-PASS "out of curiosity" or as a means of accessing a specific treatment program on AO. Also, there was a high level of mental health services and resources utilisation amongst the sample which may have reduced the need for any additional action. On the other hand, the lack of follow-up activity in some cases also pointed towards the challenges people face when dealing with mental health issues. As one individual acknowledged, they were aware of their main issues and the resources available, but found it difficult to find the right treatment and commit to action. Indeed, therapeutic change is a generally difficult, frightening, and conflicting (i.e. between conscious desires and unconscious fears) experience which naturally gives rise to ambivalence and resistance (Newman, 1994). Treatment motivation is often neglected in studies of online intervention but is posited as having a major role in the outcomes of an intervention (Bensley & Lewis, 2002). While not considered in this study, treatment motivation may be associated with the lack of follow-up response to the e-PASS in some cases.

6.4.5 Usability.

Further to Study Two, specific usability problems related to the e-PASS interface were identified. As noted in guidelines for general web design as well as design for serious mental disorders, these issues include having to scroll down certain pages (Rotondi et al., 2007) and the use of overly wordy items. Some complaints conflicted with design elements based on recommendations. For example, the e-PASS's appearance was described as "boring" by one individual, when it is generally recommended that a website's visual style not be overly stimulating (Rotondi et al., 2007). A function recommended by users was to be able to go "back" and change previous responses during the e-PASS. This function was ultimately omitted from the e-PASS due to difficulties integrating it with the complicated item branching rules. Furthermore, as mentioned by a participant, the option of revising responses might entice users to experiment with different item branching which could compromise the accuracy of e-PASS results. However, the ability to revise responses is purported to enhance a sense of anonymity and self-disclosure on computer and online instruments (e.g., Hanna et al., 2005). Also, given that some people realised they had inappropriately endorsed previous items (e.g., due to miscomprehension), the ability to change previous responses could result in more accurate e-PASS results.

6.4.6 Lengthiness.

Consistent with the findings of Study Two, a vocal complaint about the e-PASS usability was its length. As raised in other studies where computerised assessment was perceived as relatively lengthy (Greist et al., 1987; Peters et al., 1998) part of this issue may be that consumers performing autonomously can take their time to respond to items via a computer. In support of this, some participants in this study valued having more time to consider and respond to e-PASS items and "not be interrupted". However, other factors were also found to contribute to the perceived lengthiness of the e-PASS: the perceived repetition of items; the absence of a progress indicator; and the comprehensiveness of the e-PASS which differed from expectations of a briefer questionnaire (based on previous experiences of questionnaires). Although this was not statistically analysed, complaints about lengthiness also generally corresponded with actual length of completion and a greater number of e-PASS diagnoses. For example, one person who raised lengthiness as an issue took 33 minutes and received five e-PASS diagnoses. On the one hand, a longer duration is expected given the branching nature of the e-PASS which implies that people with more symptoms will likely be presented with more e-PASS items. However, it may also be that people with more psychopathology exhibit issues that extends the time it takes to complete the program (e.g., attention impairment, indecision, distress) as well as reduces their tolerance of the e-PASS length.

A notable consequence of the e-PASS's perceived lengthiness was disengagement. In contrast to Study Two, the present study revealed several accounts of careless responding as a result of this. This is a concerning finding as it suggests e-PASS performance in these cases may have been invalid, yet was not factored into the feedback provided to respondents. The difficulty in monitoring and controlling the testing environment for issues such as this is a commonly recognised disadvantage of online assessment (Barak & Buchanan, 2004). For the e-PASS, it is unclear how much impact disengagement has on the accuracy of e-PASS results. A comparison of e-PASS and CI results (from Study One) for those reporting e-PASS disengagement showed similar diagnostic results, though statistical analysis was not performed to confirm this relationship. Nevertheless, there is evidence to suggest e-PASS users are vulnerable to disengagement due to the lengthiness of the program and this may result in less attentive item responding amongst other consequences (e.g., feelings of frustration, boredom).

This study extends on the findings of Study Two by showing that those who have a need for assessment (e.g., to aid assessment with psychologist) have a greater willingness to persevere with answering items despite experiencing negative affect such as boredom and

frustration. This suggests that issues of lengthiness and disengagement are attenuated by a need for assessment addressed by the e-PASS. On the other hand, involvement with a research project may have provided additional motivation to persevere with the e-PASS, whereas others completing the e-PASS more independently might have more likely disengaged.

6.4.7 False assumptions.

Study Two suggested people generally appreciated the depth of items presented by the e-PASS. This study, however, showed a more critical response, with some people finding it annoyingly excessive and irrelevant due to the misassumptions it made particularly in relation to substance and gambling related issues. As some participants commented, the e-PASS was overly sensitive in assuming their behaviour was problematic, while not offering the opportunity to explain their circumstances. However, these participants did in fact endorse relevant screening questions which indicated substance use and gambling behaviour and therefore justified more in-depth probing. While a degree of frustration towards further questioning is understandable if substance use is not in fact problematic, the resistance expressed towards this process raises several issues that could be of importance to e-PASS use.

Given that mental disorders such as substance dependence and pathological gambling can involve a degree of defensiveness and denial (Scull & Woolcock, 2005), some people with these problems may find it challenging to have to respond to e-PASS items and receive e-PASS feedback which point towards these problems. One could imagine individuals falsely denying symptom severity either because they lack insight or deliberately want to avoid confronting the issue due to ambivalence. Interestingly, alcohol dependence was one of the few disorders which received more clinical diagnoses from the CI than the e-PASS in Study One (unfortunately, other substance disorders were not prevalent for a similar comparison). This could be evidence that web-based self-report measures such as the e-PASS struggle with identifying disorders requiring good personal insight and judgement (e.g., OCD, pathological gambling, substance disorders), and that in-depth probing of these disorders may deter individuals who lack motivation to explore these issues. However, further research is needed to explore this hypothesis.

6.4.8 The need for support.

While autonomy was a valued aspect of the e-PASS experience for various reasons (e.g., privacy, lack of judgment), it was clear that some people needed additional support either during or after the e-PASS due to aforementioned experiential issues (e.g., negative affect,

miscomprehension, lack of follow-up direction and treatment motivation). These results suggest that the e-PASS on its own does not provide an adequate assessment experience for all users, and lends support to suggestions in the literature that follow-up support be included with online testing, especially if it might elicit distress (Buchanan, 2002).

The question arises as to how support needs are best addressed. As many people may not require it, one approach could be to offer some form of therapist-assistance following the e-PASS (e.g., via email, instant messaging, phone support) as similarly employed in eTherapy interventions (Barak et al., 2008). A more restrictive approach to ensuring support is to limit the use of the e-PASS to individuals who can consult their own health professional during or after the e-PASS. A successful example of this was evident when one individual confirmed their e-PASS result with their psychologist.

However, obligating therapist-support or some form of supervision runs the risk of creating bottlenecks to assessment which online assessment programs such as the e-PASS were intended to alleviate (Christensen & Hickie, 2010). This point raises the dilemma between accessibility and ensuring best practice for consumers and service providers (Bartram, 2006). While this research did not restrict how and why the e-PASS was being used other than in the delivery model of AO, further research could evaluate the potential value of the e-PASS in alternative contexts, such as in primary care where the e-PASS could act as a preliminary screener and be completed by patients, for example, in the lead up to their next appointment.

6.4.9 Stigma of diagnostic labelling.

This study also found some negative consequences related to stigma. One provocative aspect was the warning that appeared when people endorsed having suicidal thoughts. Although recognised as a precautionary measure of screening risk and encouraging people to seek help, it left one participant feeling “freakish” and was described by others as being scary, abrupt, and overly sensitive in assigning suicidality. Concerns about the inadequacy of online testing in addressing suicidal ideation have been previously raised (Barak & English, 2002). In an exploratory study of client perceptions of online counselling, Leibert, Archer, Munson, and York (2006) found some participants described the online environment as deficient in facilitating crisis intervention. Given that roughly one in two help-seekers on the internet with a mental disorder may have suicidal ideation (Hemelrijk, van Ballegooijen, Donker, van Straten, & Kerkhof, 2012), the e-PASS is in need of a more careful and sensitive approach to addressing suicide risk.

Otherwise, the significance and frequency of stigma related experiences was less than expected of a program which made explicit diagnoses. This could be due to the e-PASS targeting higher prevalence disorders rather than more serious mental disorders (e.g., schizophrenia) strongly associated with stigma (Pitt et al., 2009). Again, a combination of prior mental health experience and disregard of the e-PASS results could also explain the lack of stigma reported by users. Of the few references to stigma, one participant spoke about how her e-PASS results may have "freaked her out" and deterred her from treatment many years ago, a comment in reference to potential disempowerment associated with clinical diagnosis (Pitt et al. 2009). However, extensive consumer experience in the mental health system had helped buffer the influence of e-PASS's results. On the other hand, this study may have missed more subtle or longer term effects of e-PASS diagnostic results, which may be especially relevant in population groups with less mental health literacy and service experience.

6.4.10 Anonymity, less judgement, and increased disclosure.

Consistent with Study Two findings, the e-PASS experience was associated with perceived anonymity and privacy due to the absence of others during testing. As expected, this corresponded with feeling less embarrassed and less judged than experienced with face-to-face assessment. These results extend on previous studies showing computerised assessment eliciting less embarrassment because of the reduced social presence compared with human administered assessment (Erdman et al., 1992; Kobak et al., 1994; Peters et al., 1998). Another factor related to performing the e-PASS autonomously was the greater sense of personal control over the assessment process. In particular, some people described feeling safer and less confronted by the e-PASS because they were not subject to the pressure of having to openly respond to potentially challenging and unpredictable questions of an interviewer. In this regard, the limited items and closed response options of the e-PASS were deemed by some as conducive in containing the assessment.

The e-PASS was also associated with promoting self-disclosure of sensitive topics. As observed in other studies, areas that were reportedly easier to disclose through a computer included suicidality (Petrie & Able, 1994) and eating issues (Winzelberg, 1997). In a powerful example, one individual acknowledged experiencing childhood sexual trauma to the e-PASS despite concealing this from previous therapists and the CI in Study One. Given that disclosing trauma is considerably difficult (Bedard-Gilligan, Echiverri, & Zoellner, 2012) yet important in the assessment and therapeutic process of PTSD, this was a significant result which demonstrates

the e-PASS's potential in eliciting very personal information. Examples of enhanced self-disclosure also reflect the notion that people are more willing to reveal their real self when online due to the online disinhibition effect amongst other factors (Barak & Hen, 2008; Suler, 2004).

While the extent of self-disclosure and its impact on the e-PASS's criterion validity is unclear across different users, the above case represents at least one instance where significant diagnostic discrepancy in Study One arose due to the disclosure of sensitive information to the e-PASS but not the CI. The result in some respects parallels the findings of Erdman et al.'s (1992) which found social desirability significantly associated with diagnostic differences between the computerised and interviewer-administered DIS. It also alludes to the limitations of clinical interviewing in eliciting sensitive information, and suggests that in some cases, a clinical interview may be inferior to web-based assessment in detecting mental disorders.

6.4.11 Constrained expression.

As observed in Study Two, there were mixed views regarding the e-PASS's response format. While some found the predominantly multiple choice items sufficient and easy to perform, others criticised it for constraining expression. Specifically, response options were seen as rigid and limited which lead people to feel as though they were compromising their responses and misrepresenting themselves. As recognised in the literature, this issue refers to a typical limitation of using close ended items in questionnaires (Schwarz, 1999). As Sheatsley (1983) describes, close ended questions force an unnatural reply and provide limited opportunity to qualify responses.

While there are benefits of using closed-ended questions (e.g., prompt responses, limit responses to categorical data, ease of response), it was evident that some e-PASS users have a need to communicate in greater and finer detail. This was especially noticeable in those who had a well formed view of their issues and had good relationships with health professional. For these individuals, questionnaires such as the e-PASS may be too simplistic in representing their understanding of their mental health issues. Also, these individuals may be more accustomed to discussing their circumstances and feeling listened to and validated. While most questionnaires employ close items, denying people the opportunity to accurately convey their issues may have important repercussions. Schulze (2007) purports that people experience a sense of stigma when their views are neglected and receive diagnostic labels that do not take into account their history and background. In this regard, the e-PASS is at risk of eliciting stigma in respondents who feel their e-PASS item responses are not reflecting their circumstances.

However, addressing the needs of those who prefer to elaborate on their responses poses a design challenge in web-based programs such as the e-PASS. One option is to incorporate more open-ended items and allow people to qualify their responses. However, this could significantly burden respondents and clinicians (i.e., analysing and scoring responses) as well as delay feedback results. While computerised software is available for processing open text responses, they currently have limited sophistication and accuracy in comparison to human judgment (Neuman, Cohen, Assaf, & Kedma, 2012). Another option could be to provide more response options to increase the likelihood of a relevant response. But this could also create more burden and confusion for respondents. Ultimately, it may be easiest to direct participants to a health professional if they are unsatisfied with how their e-PASS performance.

6.4.12 Comprehension.

Unlike in Study Two, the present study found more noticeable difficulties with item comprehension. Specifically there were complaints about questions, for instance, being overly wordy and vague. Miscomprehension was unlikely related to reading ability or education levels given the generally high level of education in the sample. Difficulties with item comprehension have been observed with traditional assessment and perhaps inevitably arise in some cases given the diverse background of respondents. For instance, Pinninti et al. (2003) surveyed respondents of the MINI interview schedule and found 11% of the sample expressed some difficulty with the structured reading of questions, while a few participants found some items too complex to understand.

However, an important aspect of the e-PASS's functionality is whether respondents can interpret and respond to the e-PASS in the intended manner. Miscomprehension and systematic response errors are two major factors that can compromise the validity of a self-report measure (Schwarz, 1999), and may explain some of the incongruence between the e-PASS and the CI. Other studies have also raised miscomprehension in explaining diagnostic differences in equivalence studies (e.g., Peters et al., 1998). It is unclear whether miscomprehension of e-PASS items did in fact compromise the validity of the results. However, difficulties of comprehending items were linked with dissatisfaction towards the e-PASS and a preference towards having someone available to clarify confusing questions, thus highlighting the significance of this issue.

6.4.13 Perceived item repetition.

Further to the findings of Study Two, the perceived repetition of e-PASS items was also raised as negative characteristic of the e-PASS experience. Interestingly, there was even suggestion that items were being deliberately repeated which exemplified how similar certain items appear for some e-PASS respondents. While DSM-IV disorders are intended to be distinct, in reality the differences can appear subtle given the symptom overlap between many disorders (Preskorn & Baker, 2002; Zbozinek et al., 2012). The perceived repetition of the e-PASS suggests that certain items may not be specific enough for participants to recognise these subtleties. This issue raises the question of whether higher rates of diagnosis for single and co-morbid disorders (especially with the anxiety disorders) seen in the e-PASS compared with the CI in Study One are an artefact of participants responding similarly to items of related disorders. In relation to this, there has been suggestion that high co-morbidity rates of common mental disorders resulting from self-report measures are due to the confusion experienced by respondents in differentiating symptoms asked of them (Slade & Watson, 2006).

It remains unclear as to how significant these issues of miscomprehension and perceived repetition are, for which items they pose an issue for, and to whom it might affect. Hence, a focus of further research could be to use cognitive testing (Schwarz & Oyserman, 2001) to investigate how respondents perceive and reply to individual items of the e-PASS. For example, a direct observation of e-PASS behaviour (e.g., to identify physical cues of item response difficulties such as hesitation) and cognitive probing of individual's thought processes as they complete the e-PASS may help in identifying and revising specific items that are more likely to create response issues (e.g., misinterpretation, provoke systematic bias, cause distress). Nevertheless, these issues highlight the inherent challenge underlying automated web-based measures such as the e-PASS of how much responsibility and judgment is left to the respondent in perceiving and responding to items.

6.4.14 Overall acceptance and preference.

This study found varying acceptance of the e-PASS. The e-PASS was often referred to as a potentially useful tool for individuals newly experiencing mental health issues to reflect on their symptoms and gain insight. For those more experienced, however, the e-PASS was not overly informative though did raise awareness of secondary issues. Despite factors such as convenience and anonymity attracting people to the e-PASS, issues such as perceived length and repetition, and difficulties comprehending and responding to items detracted from the overall

e-PASS experience. Subsequently, a number of individuals viewed undertaking an assessment with a clinician as more challenging, but ultimately more rewarding than online assessment.

These results are in line with previous research showing a preference of clinician administered assessment over computerised assessment (Erdman et al., 1992; Peters et al., 1998). They also support findings that individuals with anxiety may prefer traditional therapy approaches (e.g., exposure and cognitive therapy) over newer technologies, despite knowing treatment involves experiencing relatively high levels of stigma (Tarrier, Liversidge, & Gregg, 2006). On the other hand, this study also found some health professionals have a positive regard of the e-PASS as reflected in their referral of clients to the program. While limited, this result indicates that the e-PASS is currently being integrated with traditional assessment and supports the notion that online assessment can be used by clinicians as an adjunct intervention (Barak & English, 2002; Barak et al., 2009).

The role of the e-PASS as an online resource facilitating knowledge and insight into one's condition is similarly referred to in Lam-Po-Tang and McKay's (2010) study of psychiatric outpatients describing their general experiences of using online resources. In that study, there were a number of participants acknowledging that the low cost and potential gain in understanding their mental health issues outweighed the imperfect reliability of online information which they approached with caution. The results of the present research extend on this by showing that people are not completely trusting of online assessment, but still find trying it worthwhile because it is free and adds to their self-knowledge.

6.4.15 Limitations.

As intended, the use of qualitative interviewing was beneficial in uncovering greater depth to the e-PASS experience. In doing so, the study showed commonalities and variations in experience across a small group of people differing in personal and contextual factors. As with all qualitative research, the findings are not intended to be generalisable to the wider population. Rather, they represent the phenomena of a specific sample performing the e-PASS under certain conditions and may be transferable to similar circumstances. Although this study is argued to be relatively naturalistic compared to other user experience studies (especially usability studies where the task is performed under highly controlled conditions), there were artificial conditions in this research (e.g., knowing that one's e-PASS results would be used in the study) which may have shaped how people approached and responded to the e-PASS.

Furthermore, participants were self-selected, representing amongst other characteristics a willingness to participate in the qualitative interview. This is a beneficial for qualitative research as it implies participants more likely collaborate and share their experiences (Reid et al., 2005). However, it may have resulted in a sample in this study who were more interpersonally comfortable than typical e-PASS users. In support of this, many had previously seen or were in contact with health professionals, while some expressed a strong preference for working with a health professional because they valued interpersonal interaction. As previously mentioned, these characteristics may have influenced people's experience of the e-PASS in various ways (e.g. viewing the e-PASS as restrictive compared to consulting a health professional). While the sample still represents a subset of actual e-PASS users, the experiential themes in this study may have differed if they were drawn from those who declined participating in this study.

Although interviewing participants a period of time after e-PASS completion was advantageous in examining the short term impact of the e-PASS, the duration for some participants may have weakened their recollection of completing the e-PASS. As relevent to Study Two, the reliance on retrospective self-report in this study implies the collected data was susceptible to recall bias and may not be entirely reflective of actual experience. In the formal UE literature, it is generally agreed that actual UE is complex, situated, and temporal in nature (Hassenzahl, 2007). Furthermore, there is a difference between experience and the judgement of an experience (Hassenzahl & Tractinsky, 2006). By asking participants to recall their e-PASS experiences, this study adopted a measurement approach to experience which Hassenzahl (2007) terms as "meta-experience". This conceptualises experience as an integration of memories, where time and change are neglected, while particularly memorable aspects of experience are emphasised (e.g., strong feelings). As a result, meta-experience involves an averaging of experiences which may have been sufficient in providing an overview of the e-PASS experience in this study, but perhaps overlooks more specific data that could thoroughly inform the process of improving (i.e., through re-design) the e-PASS's UE (Hassenzahl, 2007).

Instead, monitoring e-PASS experience as it occurred may have allowed for a greater appreciation of the complex interplay between the individual, the e-PASS, the situation, and in particular, the unique and often fleeting thoughts and feelings that may arise. This could have been done by observing or interviewing participants of their UE as they undertook the e-PASS, which is an approach commonly used in usability testing (Mehlenbacher, 2002). As suggested earlier, in-person cognitive testing could be especially helpful in identifying e-PASS items that cause comprehension difficulties. On the other hand, this may have been overly intrusive for

some, therefore compromising the authenticity of participant's e-PASS experience. For example, participants may have felt more self-conscious or uncomfortable if they were observed in a foreign setting (e.g., clinic or lab) for this study. The methodological issues of how and when to measure UE is an ongoing issue in the formal UE field, in part due to the lack of specific definition around the UE construct (Hassenzahl, 2007).

Power relations may have also influenced participants' discussion of their e-PASS experience (Baxter & Eyles, 1997). As qualitative researchers highlight (Baxter & Eyles, 1997; Taylor, 2007), the status of the interviewer in areas such as class, profession, age, and gender can influence participants' reaction to an interview. In this study, the interviewer was generally of different background to participants (postgraduate research student, provisional psychologist) and had multiple roles (i.e., researcher, provisional psychologist, interviewer) when relating with participants. There was some evidence this influenced discussion as certain participants appeared keen to discuss more research related and academic aspects of the e-PASS, rather than focus on their own personal experiences. Having a familiar person conduct the UE interview had the advantage of building trust and rapport with the participant, which presumably encouraged more open and honest engagement. However, using a more neutral interviewer may have yielded more honest and less socially desirable reports of the e-PASS experience (as well as the CI in Study One).

6.5 Conclusion.

Using a qualitative approach, this study aimed to provide an in-depth and personal account of the e-PASS user experience to compliment the largely quantitative methodology employed in Study Two. This was achieved through semi-structured telephone interviewing of fifteen consumers who had recently completed the e-PASS. This study found participants generally had a history of health professional access and tried the e-PASS as a more convenient and affordable alternative to accessing traditional clinical services. Reservations were expressed towards the e-PASS's accuracy, in part due to difficulties of adequately comprehending and responding to items, and low confidence in the judgment and decision making underlying the e-PASS's branching rules. Also, as observed in Study Two, there were complaints about the length of the e-PASS and this was associated with signs of disengagement during the program.

However, the e-PASS was praised for facilitating anonymity and privacy, which in turn promoted self-disclosure of sensitive topics, and reduced feelings of judgment often associated with seeing a health professional. For one individual, this lead to an e-PASS result which was

both personally and clinically meaningful yet had been previously overlooked by health professionals. Furthermore, the e-PASS was found to promote self-reflection and raise awareness of secondary issues, and this was linked to the breadth of the e-PASS items, which participants generally valued. In general, the e-PASS was viewed as credible and valued program, however, overall acceptance was mixed with some people preferring an assessment with a health professional for the human interaction and added support, amongst other factors. Through qualitative means, this study offers a unique examination of the experience of web-based clinical assessment programs. Further research could involve different population groups and measures of usability and experience to extend on the present findings.

7.0 Chapter Seven: General Discussion and Conclusion

7.1 Overall research aim.

The main aim of this research was to evaluate a new web-based clinical diagnostic assessment program and consider its potential as a clinical assessment tool. The program, referred to as the e-PASS, was developed with the intention of producing an accurate and reliable measure of various current DSM-IV mental disorders, while offering user benefits such as convenient accessibility, ease of use, and diagnostic feedback. In general, it was hoped that the e-PASS could address barriers to adequate clinical diagnosis observed in the community (Houston et al., 2001), improve understanding and awareness of mental disorders (e.g. Jorm, 2000), and encourage access to appropriate mental health resources (e.g. online interventions). In addition to evaluating the e-PASS's psychometric properties (Study One), this research examined the user experience of the e-PASS (Studies Two and Three). These areas of investigation represent part of the e-PASS's clinical utility, or in other words, its ability to produce beneficial outcomes in real world application. The findings of this research will help promote informed and appropriate use of the e-PASS and similar programs, and also point towards areas to improve upon in future iterations. They also address the broader question of whether web-based clinical diagnostic assessment can help meet clinical assessment needs in the community.

7.2 The validity and reliability of the e-PASS.

Of primary interest when evaluating any clinical assessment instrument is whether it is valid (Summerfeldt et al., 2010). There is a paucity of studies examining both the validity and reliability of comprehensive web-based clinical assessment instruments, performed by actual users in real world conditions. Hence, Study One set out to investigate the psychometric properties of the e-PASS, focusing on criterion validity and test-retest reliability of a diagnostic result. Participants first completed the e-PASS, with subsets going on to complete a telephone administered structured clinical interview (CI), an online questionnaire consisting of various standardised clinical scales, and e-PASS retesting. Despite the small sample size (n = 39), the e-PASS displayed promisingly high test-retest reliability across a number of disorders and over a mean of eight days between test and retest. While these results need confirmation with a larger sample, they suggest the e-PASS can produce generally reliable diagnostic outcomes, which is an essential pre-requisite of validity.

Of main psychometric focus was the e-PASS's criterion validity. With the diagnostic results of the CI as the primary validity criterion, the e-PASS demonstrated generally excellent specificity but varying sensitivity in classifying a clinical disorder. Overall, there was high diagnostic agreement between the e-PASS and CI for some disorders (e.g. panic disorder) but not others (e.g. OCD, social phobia). Some of the observed disagreement could be explained by differences in the diagnostic threshold of the e-PASS and the CI. Consistent with this, reducing the e-PASS's severity threshold (i.e. at which reported distress and interference warrants a clinical disorder) was found to help balance the e-PASS's ability to rule in and out a clinical disorder, particularly in the case of OCD and social phobia. Furthermore, results indicated that adding certain items from other questionnaires could enhance the diagnostic accuracy of the e-PASS. An e-PASS clinical diagnostic result was also found to have a significant effect on the scores of relevant clinical questionnaires (e.g. CES-D, SPIN, and Y-BOCS-SR) and the K-6 measure of distress and impairment, with moderate to large effect sizes. These results were indicative of the e-PASS's convergent validity. Finally, the e-PASS produced higher rates of clinical disorders and co-morbidity than the CI, while displaying mostly poor to moderate agreement with the CI in identifying the primary diagnosis.

7.3 The user experience of the e-PASS.

Further evaluation looked at how e-PASS users perceived their overall experience of the program. In Study Two, an online survey was developed to measure people's view of their e-PASS experience in comparison to more traditional forms of clinical assessment (e.g. clinical interview, consultation with GP). There was general support for the overall acceptance and usability of the e-PASS, with many users highlighting the convenience, anonymity, ease of use, comprehensiveness of assessment, and provision of feedback as positive aspects of completing the e-PASS. However, there were also e-PASS characteristics detracting from the experience, such as the perceived lengthiness and impersonal nature of the program, and the ambiguity and subsequent miscomprehension of items. In particular, the e-PASS was associated with difficulty in expressing oneself and this greatly contrasted with the experience of the CI in Study One. In general, individuals endorsing a need for assessment displayed better engagement and more positive outcomes (e.g. greater motivation to seeking consultation with a health professional) when completing the e-PASS. Certain aspects of the e-PASS experience were also found to predict the diagnostic disagreement observed in Study One, though causality could not be clearly determined.

In Study Three, qualitative interviewing of a small sample of e-PASS users provided a closer examination of e-PASS experience. This reflected diverse e-PASS experiences with varying motivations for use, interaction issues, and outcomes of performing the e-PASS. Interestingly, some users commenced the e-PASS with significant mental health service experience and an established view of their presenting issues, yet still benefitted from the self-reflection and increased awareness of previously unrecognised problems elicited by the e-PASS. Performance issues such as disengagement and miscomprehension were again apparent, and in some cases, suggested a direct threat to e-PASS validity, though of uncertain extent.

A desire for additional support was also expressed by some individuals wanting item clarification, emotional support, and guidance at the program's end. However, the absence of an assessor/interviewer was widely valued for offering a sense of autonomy, anonymity, and privacy which provided a safer and less judgmental space for people to disclose more personal issues such as trauma, suicidal ideation, and disordered eating. For one participant, this culminated in an e-PASS diagnosis which was personally meaningful for the individual and clinically significant in representing their symptoms and informing appropriate treatment. This case also exemplified how the e-PASS, in some instances, was more accurate than the CI in Study One, and that associated diagnostic disagreement reflected the superiority of the e-PASS over the CI in eliciting more honest responses.

7.4 Application of the e-PASS.

As the e-PASS has relative strengths and weaknesses, the appropriateness of its use will likely depend on the requirements of the situation. When a highly accurate diagnosis is required in a "high stakes" situation (i.e. when diagnosis has significant repercussions; Bartram, 2006), it is clearly inappropriate to solely rely on the e-PASS given its variable psychometric performance across different disorders and its susceptibility to invalid completion (e.g., due to miscomprehension, reckless responding, fatigue). Also, it is important to remember the e-PASS has a number of diagnostic limitations due to the nature of the program. Specifically, it is focused on current diagnosis, certain disorders, and is limited in factoring contextual information. Although the e-PASS screens symptom onset and the causal role of medical conditions, substance use, and other issues (e.g. bereavement, previous diagnoses), this is insufficient in more complex presentations (e.g. when medical conditions cause anxiety; depression during bipolar disorder) which ideally require a full assessment of history, additional testing, and clinical/medical expertise to establish diagnosis.

As seen in Study Three, e-PASS users who experience and recognise having complex symptom presentations may experience frustration in not being able to fully clarify this with the e-PASS and subsequently doubt the accuracy of their results. In such cases, a more thorough clinical assessment with a mental health professional would be ideal in allowing the consumer to openly discuss their concerns, as well as confirming diagnosis. While e-PASS users are already warned that the e-PASS does not replace a formal clinical assessment and encourages users to undertake this with a health professional, a more thorough statement about the program's main psychometric characteristics and user experience issues as detailed in this research could help promote informed consent and appropriate use.

Despite its shortcomings, the e-PASS still offers diagnostic utility, particularly for health professionals. Meyer et al. (2001) suggested that clinicians who rely on one source of assessment (e.g., clinical interviewing) are at risk of neglecting important assessment information. Therefore it is suggested that clinicians incorporate other assessment means such as routine testing or brief questionnaires to maximise data collection (Meyer et al., 2001). In this regard, web-based clinical questionnaires such as the e-PASS, which offer a range of practical benefits over other forms of clinical assessment (e.g. paper-pencil questionnaire), could be very suitable as a diagnostic aid to compliment clinicians' usual assessment practice. As evident in Study Three, this practice is already occurring, with some professionals referring clients to complete the e-PASS during or in between sessions, then subsequently discussing results and conducting further assessment where needed.

One key area where the e-PASS is likely to be beneficial is in assessing numerous disorders, and especially lower prevalence disorders. As noted in Study Two, the e-PASS is reportedly more comprehensive than community-based assessments, including by health professionals, and this is likely because it resembles a structured interview schedule (e.g. MINI-Plus) in covering a range of disorders. Although structured interviewing has been shown to identify up to five times more diagnoses than the typically unstructured approach conducted by clinicians (Ramirez-Basco et al., 2000), it is overlooked in clinical practice because of constraints in time and resources. Therefore, it may be advantageous for clinicians to employ the e-PASS as a multi-disorder screening tool without having to manually administer a structured clinical interview schedule. This could make better use of consultation time and help identify co-morbidity. However, as exclusionary criteria are not factored into e-PASS diagnostic results, some co-morbidity rates might be higher than what is actually present, therefore requiring clinical judgment in determining whether there are redundant diagnoses.

Nevertheless, the e-PASS can screen disorders with reasonable accuracy. As found in Study One, the e-PASS recognises a large proportion of actual clinical disorders through its subclinical and clinical diagnostic results, both of which clients and clinicians would presumably act upon. Also, the e-PASS is relatively conservative in assigning a clinical disorder amongst those who do not actually present one, and is generally correct when ruling out a clinical disorder. Referring to predictive values, one can be moderately confident that a clinical e-PASS result will actually represent a clinical disorder (or possibly even a subclinical presentation, though this was not investigated in Study One), and very confident that a non-clinical e-PASS result will reflect the absence of one. These characteristics differ from many publically available online screening programs which tend to have low thresholds and produce a high level of false-positive results and could result in unnecessary burden (e.g. further assessment).

There are also several experiential factors that should be considered by clinicians and consumers when using the e-PASS. Given that the e-PASS is heavily text-based, those with poor language ability may find it difficult comprehending and responding to items. Furthermore, the e-PASS is lengthy compared to typical online screening tools and paper-pencil scales, which may result in disengagement, particularly in individuals with low assessment motivation. For those inexperienced with web-based questionnaires, it is also worth highlighting that the e-PASS lacks a personable experience due to its automated nature and lack of human presence. Across the user experience studies, this was evident in criticisms of the e-PASS being “cold” and restrictive in response options, and not offering adequate support for issues such as miscomprehension, distress, and low treatment motivation upon completion. For some people, these difficulties tainted their e-PASS experience and reinforced their preference for seeing a health professional. Such experiential issues echo Naglieri et al.’s (2004) assertion that the standardised and often impersonal nature of online testing may not be appropriate for everyone.

For most e-PASS users, however, there are important benefits of being able to perform the e-PASS on one’s own. In particular, the e-PASS can provide a welcomed sense of anonymity and privacy which is free of the judgment, pressure, and associated feelings of discomfort that can arise when asked personal questions by a health professional. A consequence of this is the possibility that the e-PASS will elicit diagnostic information that may not as freely arise in a clinical interview, as reported in Study Three. However, participants’ willingness to reveal sensitive information afforded by the online disinhibition effect may be constricted by the limited items and response options of the e-PASS. As observed in Study Three, this could be met with frustration by some people wanting to discuss their mental health issues in more detail, whereas others

might appreciate that the e-PASS avoids delving too deeply into one's mental health issues. Regardless, the e-PASS could be a valuable tool for encouraging consumers to contemplate aspects of their mental health. This could also be particularly helpful for clinicians in eliciting sensitive, yet diagnostically significant information from clients referred to the e-PASS, though care would be needed in following up on this (e.g. addressing reported suicidality or trauma).

Many of the experiential issues associated with the e-PASS are shared with other forms of clinical assessment (e.g. limited response options in paper-pencil questionnaires, lengthiness of structured interview schedules) or inherent in the typically autonomous context of undertaking an online program. Nevertheless, there is a need to inform potential users about these limitations, and screen for those who may be particularly affected by these experiential issues (Naglieri et al., 2004), to help mitigate negative experiences (e.g. dissatisfaction, disengagement) and ensure appropriate e-PASS feedback. For example, regarding the length of the e-PASS, users at the beginning of the e-PASS could be clearly notified about how greater symptom endorsement results in lengthier administration, and the use of breaks when feeling tired during the program.

With clearer expectations, e-PASS users may be more compliant and may implement strategies to maintain engagement, though warnings of lengthy administration times could also deter those preferring a quick test. Another option could be to screen for those who might adversely react to the length of the e-PASS, and to either offer them a shorter online instrument such as the WSQ (Donker et al., 2008) or offer a staggered version of the e-PASS. Screening could, for example, ask about attentional difficulties or, as indicated in Study Two, target people with depression who may be less tolerant of the e-PASS's length. Finally, screening of experiential issues (e.g. difficulties with item comprehension) following e-PASS completion could indicate the validity of e-PASS results and assist appropriate interpretation. While these various measures of addressing experiential issues could be included in the e-PASS program, in the meantime, health professionals could directly communicate them with clients using the e-PASS.

7.5 Stepped approach to clinical assessment.

Roughly half of the e-PASS users in this research were managing their mental health on their own, with many intending to access self-help treatment. Although the e-PASS may not substitute a formal clinical assessment, an important question is whether it may still sufficiently inform appropriate treatment. The e-PASS is situated in a system which directly links diagnostic assessment results with online treatment programs (Klein et al., 2011, 2012). Implicit in this design is the notion that brief assessment (i.e. the e-PASS) is adequate for brief intervention

(Baillie & Rapee, 2004). This approach resembles a stepped care model of treatment and intervention, except in the assessment context the process also begins with a low intensive form of assessment. If the resulting (or lack of) intervention does not lead to recovery, then a more detailed assessment and/or alternative intervention is warranted (Baillie & Rapee, 2004).

A stepped approach contrasts to the ideal practice of firstly conducting a detailed assessment to choose the intervention with the greatest chance of success (Newman, 2000). However, as Baillie and Rapee (2004) highlights, a detailed clinical assessment can be burdensome and, across different circumstances, will likely result in the same treatment recommendations given the few empirically validated treatment approaches to choose from and their limited specificity (e.g. certain medications and psychotherapeutic approaches can be effective across various disorders). Hence, it is argued that a detailed clinical assessment can often produce superfluous information with little practical benefit in informing appropriate intervention (Baillie & Rapee, 2004).

The concept of coupling brief assessment and intervention underlies the intended role of the e-PASS at AO, and is similarly referred to by researchers advocating the use of brief online screening instruments to help refer people to appropriate online intervention programs without needing more intensive assessment (Christensen & Hickie, 2010; Donker et al., 2009). There is evidence demonstrating the efficacy of the approach in relation to the e-PASS. Specifically, Klein et al. (2011, 2012) found that e-PASS users who went onto complete recommended online treatment programs based on e-PASS results showed a reduction in symptom severity according to the outcome measures (i.e. e-PASS scores and K-6). However, further research is needed to compare the clinical utility and longer term outcomes of employing brief assessment such as the e-PASS, against a more thorough clinical assessment procedure such as a structured interview by a clinical psychologist or psychiatrist. While the latter would undoubtedly involve a more thorough assessment, it is unclear whether it ultimately produces better outcomes (e.g. clinical, financial) for an individual.

7.6 Treatment measure.

An important function of clinical assessment is to measure treatment outcomes (Hunsley & Mash, 2010), and, consistent with this, the e-PASS is used for pre- and post- evaluation of online intervention outcomes in the AO system (Klein et al., 2011, 2012). The present research evaluated diagnostic accuracy but did not consider whether the e-PASS could accurately reflect diagnostic changes as a result of treatment. In the aforementioned studies by Klein et al. (2011,

2012), where the e-PASS was used as one of several outcome measures, treatment related changes in e-PASS results were broadly consistent with those of the K-6 and a quality-of-life item, though associations were not statistically analysed. While this provides some evidence of the e-PASS's validity as a treatment measure, further research could directly compare the e-PASS's performance against changes between pre- and post-treatment outcomes of more robust validity criteria (e.g. structured interview, established symptom scales).

A related matter is whether e-PASS severity scores could and/or should be used instead of diagnostic results (i.e. clinical vs. non-clinical disorder) as a treatment measure. In psychiatric research (e.g. medication trials), dichotomous outcomes have been traditionally used in determining recovery and remission. However, continuous measures have been generally preferred by psychotherapy researchers (Cuijpers et al., 2010a), likely due to having greater interest in gradual changes in symptoms, severity, and underlying dimensional constructs of psychopathology. Although a clinical result in the e-PASS is based on a threshold applied to severity scores, different treatment effect sizes could result when using a dichotomous outcome instead of the continuous measure it is based upon (Cuijpers et al., 2010a).

Furthermore, it is uncertain how the e-PASS would perform in measuring latent constructs. Compared with standardised clinical scales such as those employed in Study One (e.g. CES-D), the e-PASS shares some item content and was found to have similar if not slightly better diagnostic accuracy compared to some scales. Also, an e-PASS clinical result appropriately corresponded with a significant effect on scale scores. However, the e-PASS has not been evaluated for construct validity. While it is assumed that e-PASS items based on DSM-IV diagnostic criteria represent the intended disorders, latent trait analysis could examine the relationship between e-PASS items and their relevance to mental disorder constructs. This would require administering all e-PASS items (i.e. without branching rules) or employing a very large sample to ensure all e-PASS items have sufficient data for analysis. Analysis of construct validity could bypass the limitations of criterion validity (i.e. the lack of a gold-standard) and inform the true pattern of symptoms measured by the e-PASS. It could also help clarify the relationship of e-PASS items and help explain patterns of co-morbidity between e-PASS disorders.

7.7 Impact.

This research showed that the e-PASS enhanced awareness of mental health issues and had an effect on follow-up behaviour for a portion of users. For some, the e-PASS prompts motivation to consult with a health professional (especially if there is an existing need for

assessment and/or relationship with a health professional) which can result in a significant intervention in terms of guiding further assessment and treatment, and providing personal benefits associated with receiving a meaningful diagnosis. For many, however, the e-PASS had little influence and this was linked to prior awareness of issues, mental health experience, and problems with using the e-PASS. While Study Two employed a larger sample which could better represent e-PASS users, it did not directly investigate follow-up behaviour. However, it did suggest that on average, the e-PASS encouraged people to speak to a health professional, though to a lesser extent than the CI.

The varying impact of the e-PASS amongst participants reflects the complexity underlying whether an online assessment program such as the e-PASS can lead to positive outcomes. As theoretical models proposed by Ritterband et al. (2009) and Chiu and Eysenback (2010) outline, there are a multitude of factors that could influence behaviour change. The results of this study reflect some of these factors, such as the influence of support, the website characteristics (e.g., length), and motivation. However, a closer examination of how these factors directly impact on the e-PASS outcomes is needed. While recent research has begun examining the follow-up outcomes of web-based screening (Parker et al., 2013), there is still limited understanding on the longer term impact of online assessment. Hence, further research could investigate how people are impacted by their e-PASS results and recommendations at different follow-up points (e.g., three, six, and nine months after completing the e-PASS; Van Ameringen et al., 2010). A randomised control trial could also investigate the effect of the e-PASS against conditions such as a waiting group and brief in-person assessment. This could reveal important online assessment outcomes such as the extent to which online results influence follow-up emotions, beliefs, and behaviour.

7.8 Cost-effectiveness.

Related to outcomes is the issue of cost-effectiveness. An often noted advantage of web-based assessment over traditional methods is the lower incremental cost of delivery (Naglieri et al., 2004). While specific e-PASS costs are unknown, the initial development cost (AUD$1.6 million) and ongoing costs of running the entire AO system (including the e-PASS, the treatment programs and other services) during the first 18 months of operations totaled AUD$2 million (Klein et al., 2011). During a similar period, 9,085 people completed the e-PASS. Based on recommended psychologist fees for clinical assessment (i.e., AUD$222 for 46-60 minutes; APS, 2013), and the average clinical interviewing time in this research (i.e., *M* = 48 minutes), delivering

similar diagnostic assessment by a psychologist to 9,085 people would cost a total of AUD$2,016,870. Therefore, one could infer that the e-PASS offered a saving of approximately AUD$17,000 before even factoring the savings related to the treatment programs offered by the rest of the AO system which, according to Klein et al. (2011), would bring the total savings to approximately AUD$6.7 million. The incremental savings would be even greater over time given that initial development costs are not recurring (Klein et al., 2011) and the e-PASS continues to be highly accessed by the public.

Naturally, a freely available program such as the e-PASS provides a considerable saving for consumers compared to seeing a health professional for a diagnostic assessment. As commented in Study Three, some people are not in a "position to pay a psychologist $130 per week", and therefore, the e-PASS offers a financially appealing alternative to accessing traditional diagnostic assessment. However, an examination of the e-PASS's cost-effectiveness, from a service delivery and consumer perspective, needs to also consider the longer term costs associated with diagnostic results, follow-up actions, and outcomes (e.g. inaction, further assessment, treatment). For example, a misdiagnosis may be costly for an individual if it leads to unnecessary or ineffective consultation and treatment.

It is important to note that a program such as the e-PASS may not have to be as clinically effective as a gold-standard form of assessment to demonstrate overall cost-effectiveness. From a population perspective, for instance, the cost of misdiagnosis in some cases may be offset by the overall savings of correctly diagnosing a large proportion of the many people who perform the e-PASS. Few studies have thoroughly examined the cost-effectiveness of online interventions (Tate, Finkelstein, Khavjou, & Gustafson, 2009), let alone web-based questionnaires. A study by Smits, Smit, Cuijpers, and De Graaf (2007) showed that adjusting a cut-off score for a screening instrument (General Health Questionnaire) used to screen depression could result in varying costs in different contexts (e.g. research versus consumer versus health service provider). A thorough investigation of the cost-effectiveness of the e-PASS could help clarify its economic value, and inform policymakers and stakeholders as to whether to prioritise investment in online programs such as the e-PASS.

7.9 Future challenges and opportunities.

7.9.1 DSM and the future of psychiatric classification.

There are several notable challenges to the longer term application of the e-PASS. While it will likely face initial resistance, the adoption of DSM-5 will eventually render assessment

instruments based on DSM-IV obsolete. Hence, the e-PASS will need to be revised in line with the criteria outlined in DSM-5 if it is to remain relevant. An inspection of the recently published DSM-5 (www.dsm5.org, December 2012) suggests these proposed changes will more likely be minor than major. Nevertheless, once revised in accordance with these changes, the e-PASS will need to undergo further evaluation of its psychometric properties and user experience. Regarding criterion validity, a short-term obstacle to evaluating this will be finding a suitable criterion given it may be a while before new or revised measures based on DSM-5 are available and psychometrically established. Presumably, DSM-5 will be associated with improved diagnostic criteria. Hence these changes may enhance the e-PASS's psychometric properties as previously seen in diagnostic instruments that were revised from DSM-III to DSM-IV (Brown et al., 2001).

Another challenge is the possibility that psychiatric assessment evolves to incorporate a more dimensional approach to formally representing psychopathology. While a largely categorical model will likely continue to underlie psychiatric nomenclature (i.e., as evident in DSM-5) in the near future, there is growing evidence in support of dimensional models in describing how mental disorders actually exist (e.g. Krueger, McGue, & Iacano, 2001; Krueger & Markon, 2006; Naragon-Gainey & Watson, 2011). Krueger and colleagues (2001), for instance, found a higher-order structure of common mental disorders consisting of two factors ("internalising" and "externalising") which in turn correlated with personality traits. There is mounting evidence that these broad factors accurately account for co-morbidity and explain shared characteristics (e.g. genetic factors, biological processes) between many common disorders (Krueger & Markon, 2006; Ofrat & Krueger, 2012).

As the dimensional literature builds momentum, computer adaptive testing (CAT) based on item response theory could become more relevant in web-based clinical assessment. As referred to in Chapter Two, CAT is ideal for measuring latent traits while offering practical benefits of reducing response burden and floor or ceiling effects compared to measures based on classical test theory (Butcher, 2003; Fliege et al., 2005). A CAT program could, for example, be used to efficiently identify where people are positioned on the internalising and externalising dimensions previously mentioned, and this could then indicate information such as the likely disorder, risk factors, and appropriate treatment. However, CAT faces several practical challenges (Walter et al., 2007). Also, further research in the dimensionality of psychopathology (e.g. what are the broad dimensions and how do DSM disorders map onto these?) is needed before it can provide sufficient clinical utility for the purposes of clinical diagnosis/assessment.

7.9.2 Idiopathic assessment.

One diagnostic issue related to the e-PASS and web-based assessment programs in general is that certain aspects are applied nomothetically or generally across a range of people (Buchanan, 2003). For example, to reach diagnostic threshold, e-PASS scores are measured and compared against a group reference, which at present is a severity threshold of 3.5. However, a problem with this approach is that a standard threshold may be generally appropriate for a population group, but may not be appropriate at an individual level given the variation that can arise from one individual to the next (Sale, 2006). This issue is exacerbated for online programs such as the e-PASS given the wide range of people who could potentially access the program.

In light of this, perhaps a more appropriate application of the e-PASS consistent with the general aim of assessment involves an idiographic approach which focuses on the individual (Buchanan, 2003). For instance, rather than compare an individual's severity score against a group-normalised diagnostic threshold to determine clinical significance, the e-PASS could remove the threshold and simply reflect the profile of key symptoms endorsed by the user and encourage decision making based on this, and ongoing e-PASS assessment as a means of tracking symptom changes. This could have several benefits: it bypasses the issue of obtaining appropriate normative data (Buchanan, 2003); it focuses on symptoms rather than the overall diagnostic result which could be more acceptable for consumers; and it recognises the complexity and idiosyncrasy of individual symptom profiles, leaving responsibility for clinicians in formally diagnosing any underlying mental disorders.

7.9.3 Client-centred assessment.

While brief diagnostic assessment may have its advantages, one could argue there is an opportunity to expand the role of the e-PASS so that it is more centered on the needs and circumstances of the individual, rather than being focused purely on diagnosis and catering for the "average user". As Harwood et al. (2011) highlights, there is a risk that online applications oversimplify the assessment and treatment process, and that important information is overlooked because of technological limitations and an over-reliance on the model of fitting treatment to disorder diagnosis. In contrast, an example of a person-centred program is the Systematic Treatment Selection (STS) system available at www.innerlife.com (Harwood et al., 2011).

The STS system involves a web-based questionnaire which assesses personal variables and dimensions that, according to STS principles, predict outcomes with different psychological interventions. These factors include symptoms, coping style, social support, subjective distress,

functional impairment, trait-like resistance, stages of change (readiness), amongst others (Beutler, Alomohamed, Moleiro, & Romanelli, 2002; Harwood et al., 2011). STS results follow a cross-cutting approach of matching a person's attributes with interventions and therapeutic strategies (e.g. recommended therapist style) that are expected to optimise treatment outcomes. Although the STS system has not been evaluated, the empirical and theoretical evidence of its underlying principles support its efficacy (Harwood et al., 2011).

At present, the e-PASS also collects a range of information such as motivation and preferences for learning which could influence treatment interaction and outcomes. However, e-PASS feedback results and referral recommendations do not currently incorporate this information. As the STS exemplifies, there is perhaps an opportunity for web-based programs to assess and provide feedback on a range of factors which could either help match users with appropriate treatment and intervention, or inform how interventions (especially online programs) could be tailored to the individual. However, as mentioned earlier such efforts require consideration of whether additional assessment is in fact necessary and whether it will result in better outcomes compared to a briefer form of assessment. Inevitably, there will be some situations where this additional assessment will be valued by consumers and justified by improved outcomes, while others may find it excessive and ultimately unhelpful.

7.9.4 Incorporating multimedia.

Due to constraints and limited innovation in the development of new diagnostic instruments (Butcher, 2006), many current web-based programs (including the e-PASS) resemble traditional questionnaires and interview schedules in relying on the text-based questionnaire format. While this approach is likely familiar to users, it underutilises the opportunity to incorporate multimedia as increasingly evident in other online resources (e.g. CBT treatment programs featuring videos and diagrams). Barak and Buchanan (2004) earlier proposed the use of the web in facilitating projective testing (e.g., Rorschach inkblot or Thematic Apperception Test). However, there has since been little empirical evidence of such applications, likely related to the declining status of projective tests and their demand on clinician expertise.

A more practical approach could be to supplement text-based web assessment by incorporating images and/or audio (Chinman et al., 2007; van Ballegooijen et al., 2012). Van Ballegooijen et al. (2012) examined the web-administration of the Visual Screener for Common Mental Disorders, and focused on the validity of a single item agoraphobia measure which visually depicts four crowded situations and is accompanied by a text and audio presented

question about the fear of such situations. With the CIDI as criteria, the overall diagnostic accuracy of the program in screening for agoraphobia was found to be moderate, suggesting promise with further development (van Ballegooijen et al., 2012). The e-PASS could similarly be adapted to include at least audio presentations of items to aid engagement and comprehension, particularly for those with limited reading or visual ability.

7.9.5 Cultural sensitivity.

In tailoring to the needs of the user, web-based programs could especially focus on being more culturally sensitive. There is increasing diversity across populations (e.g. Australia, US) implying the need for clinical assessment instruments which are appropriate for different cultural groups (Butcher, 2006). Furthermore, given the relatively open access of programs such as the e-PASS, users could have diverse cultural backgrounds (e.g. accessing from different countries) which might influence their experience of the program and subsequent results. Online programs seem ideal in tailoring to different cultural groups, for example, by automatically offering translated text or accompanying audio when the user is identified as preferring another language. This offers added utility for clinicians (e.g. primary care) working with diverse client backgrounds. Accordingly, the e-PASS could be adapted to cater for other languages and cultures, though ensuring the appropriateness of the translation and diagnostic validity may be challenging.

7.9.6 Behavioural and situational testing.

There are also exciting assessment opportunities arising through the observation of behavior in an online environment. The online space can enable a high level of presence, especially in highly immersive environments, and encourages people to act more freely and openly by providing a sense of anonymity (Barak & Hen, 2008). Online behaviour can therefore reveal richer diagnostic information compared with formal clinical interviewing or self-report questionnaires such as the e-PASS. For example, peer-support forums and blogs can allow the indirect assessment of clinical issues (e.g. interpersonal issues, suicidal behaviour) expressed through personal expression and social interaction (Barak & Buchanan, 2004). Neuman et al. (2012) for example recently developed a system which scanned web blogs and classified depression symptoms with 84.2% accuracy compared with clinician judgment. Such technology could also be incorporated into web-based questionnaires as well as in synchronous text based assessment methods (instant messaging or email) as a way of screening for symptoms expressed through open text.

Clinicians could also assess how individuals participate in virtual environments designed to elicit affective, cognitive, and behavioural responses. Earlier research by Riva (1998) investigated the use of a computer program (also available online) which presented 3D images of different shaped figures to assess body image. Subjects chose the figure that they felt best reflected their current and ideal body sizes, and the discrepancy between this reflected current dissatisfaction. There has also been extensive research in the use of more immersive virtual reality technology (e.g. head mounted display) for neuropsychological assessment and exposure therapy of anxiety disorders that have shown efficacy (Kang et al., 2008; Rizzo, Parsons, Kenny, & Buckwalter, 2012). While these technologies have previously been costly and cumbersome, it appears similar functionality will soon become more accessible as personal online computing devices offer increasing immersion into a virtual (e.g. Second Life) or augmented reality (e.g. Google Glass).

7.9.7 Mobile delivery.

One exciting medium for delivering clinical assessment involves mobile technology. As the internet becomes increasingly mobile, there is a consumer shift away from desktop computers to portable online devices (e.g. phones, tablets, wearable devices) which could facilitate assessment in the moment and in everyday situations (Wichers et al., 2011). Unlike previous portable computing devices (e.g. PDAs, laptops) used for assessment, current mobile devices allow for several functional advancements. Modern smart phones for example have a range of sensors (e.g. location, movement, biometric, audio, and visual) that could broaden data collection methods while reducing reliance on retrospective self-report information. Mobile internet access maintains instantaneous data communication, for example, to health professionals who could be alerted of significant clinical issues detected by the device. As mobile devices are usually left running and kept at hand throughout the day, they offer a highly dynamic, timely, and convenient source of assessment data.

In the medical and health community, mobile devices have long been utilised to collect both self-reported and neurophysiological information (Free, Phillips, et al., 2013). Referred to as mobile health or "mHealth", this approach has more recently been applied to clinical assessment, and particularly for monitoring mental health-related information. For example, researchers have developed and evaluated the use of mobile phone applications to help track young people's mood, stress, and coping (Kauer, Reid, Sanci, & Patton, 2009; Reid et al., 2009). Reid et al. (2009) investigated the feasibility of using mobile phones in administering brief questionnaires for

monitoring activity, mood, stresses, and substance use. Their program ("mobiletype) prompted questionnaire responses several times a day, and sent the data via SMS to researchers. Reid et al. found that the mobiletype elicited a high level of engagement and endorsement of accuracy, leading to suggestion that mobile phones could be especially helpful in promoting self-awareness of emotions and stressors, which in turn could facilitate more adaptive coping strategies.

Commercial smart phone applications for screening of mental disorders are also becoming available to consumers and clinicians. A recent example is the "WhatsMyM3" app available on the IOS and Android mobile platform, based on the My Mood Monitor questionnaire of Depressive, Bipolar, Anxiety, and Post-traumatic Stress Disorders for use in primary care (www.whatsmym3.com). While the app is yet to be evaluated, the original paper-pencil checklist displayed good sensitivity and specificity when compared with the MINI (Gaynes et al., 2010).

The e-PASS could similarly be converted into a mobile application or optimised for web-use on a mobile device, though the added utility of this over the current delivery via desktop web-browser is uncertain. On the one hand, the novelty of downloading an application and storing it on one's personal device may encourage accessibility and greater (i.e. repeated) use of the e-PASS, especially if the application offers functionality such as storage of previous responses and results, to allow for personal tracking of one's mental health. However, it is unclear whether people would want to undertake a lengthy and comprehensive questionnaire requiring extended attention on a smaller mobile device, and whether the flexibility of mobile use would increase the risk of inappropriate assessment behaviour and conditions (e.g. performing on a busy train).

7.9.8 Integrating sensors.

The integration of sensors in computers and mobile devices (e.g. microphone, camera) could play an important role in future clinical assessment. Researchers have begun identifying visual and auditory cues of certain disorders and looking at ways of integrating this with computer technology for the purposes of assessment (Cohn et al., 2009; Joshi, Dhall, Goecke, Breakspear, & Parker, 2012). For instance, Joshi et al. developed a computerised system focusing on the spatio-temporal features and changes of upper body gestures and facial expressions associated with depression. In a clinical trial where video clips of people with and without depression (as diagnosed by two independent clinicians) were analysed by the computerised program, a maximum accuracy of 88.3% was achieved.

Given these promising results, Goecke and colleagues are in the process of developing a multimodal program for diagnosing depression which examines a combination of facial

movements, expressions, and a person's voice (http://ise.canberra.edu.au/roland/news/). This technology effectively computerises the process that clinicians undertake in assessing the non-verbal aspects of depressed behaviour. While Groecke and colleagues intend to develop a mobile application for consumer use in monitoring depression, one could imagine this technology integrated into forms of telepsychiatry (e.g. videoconferencing) to provide an additional diagnostic tool for clinical interviewing. An alternative approach that omits the involvement of an assessor could be to supplement this multi-modal tool with web-based questionnaires such as the e-PASS.

7.9.9 Online clinical assessment centre.

This research focused on one specific form of online clinical assessment, namely the web-based questionnaire. However, there are other commonly used forms, such as clinical interviewing via videoconferencing or instant messaging, which offer relative advantages and disadvantages. There are no studies known to this author which have empirically compared the performance of web-based questionnaires with these other forms. However, rather than preferencing one over another, a more clinically rigourous approach could be combining them (as well as upcoming forms of online assessment) to leverage their respective strengths and provide a more comprehensive clinical assessment (Barak & Hen, 2008). Barak and Hen (2008) referred to this when proposing the concept of online psychological assessment centers which integrate online interviewing, testing, and situational tasks for purposes such as recruitment and training in organisational contexts. This could extend to clinical assessment purposes and help maximise sources of data collection as recommended for enhancing validity (Meyer et al., 2001).

7.10 Final conclusion.

As highlighted in Chapter One and apparent throughout this research, mental disorders cause enormous impact across the community. Effective clinical assessment is necessary to help distinguish mental disorders and guide appropriate treatment. The internet has offered new ways for delivering clinical diagnostic assessment. To encourage the appropriate use of such assessments, it is important that efforts be made to clarify associated characteristics, including advantages and limitations (Barak & English, 2002). Hence, this research set out to evaluate the Online Psychological Assessment (e-PASS) program, a newly developed web-based clinical diagnostic assessment tool which targets a range of disorders. This research identified relative advantages and disadvantages of the e-PASS compared to traditional forms of assessments, and these qualities could extend to other forms of web-based clinical diagnostic assessment programs. Importantly, the e-PASS can exhibit a high level of reliability and diagnostic agreement

with a structured clinical interview and other standardised clinical measures. It is also acceptable to users, offering benefits such as affordability, convenience, anonymity, and access to a comprehensive form of assessment which, for many people, outweigh the limitations of the e-PASS and its web-based questionnaire format.

Whether it is used by consumers who cannot or are unwilling to see a health professional, or by clinicians requiring a practical diagnostic aid, the e-PASS and similar web-based programs have the potential to facilitate clinical assessment and result in positive outcomes. Further to its primary role in identifying mental disorders, the e-PASS can act as an intervention, prompting self-reflection and improved mental health awareness and literacy. In particular, the e-PASS offers consumers a means of overcoming the bottle-neck associated with traditional clinical assessment, allowing more immediate and well-informed access to specific online interventions. Further research is needed to clarify the findings of this research, and to demonstrate the effectiveness of the e-PASS under different circumstances before it can be strongly endorsed and widely adopted. However, this research and its findings support the vocal assertion and growing evidence that online resources including web-based clinical assessment programs can help address the impact of mental disorders.

References

Abbar, M. (1996). Panic disorder and panic attack. *Encephale, 22*(5), 13-18.

Aboraya, A., France, C., Young, J., Curci, K., & LePage, J. (2005). The validity of psychiatric diagnosis revisited. *Psychiatry 2*(9), 48-55.

Adams, N., Stubbs, D., & Woods, V. (2005). Psychological barriers to internet usage among older adults in the UK. *Medical Informatics and the Internet in Medicine, 30*, 3-17.

Alarcon, R. D. (2009). Culture, cultural factors and psychiatric diagnosis: Review and projections. *World Psychiatry, 8*, 131-139.

Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism, Clinical and Experimental Research, 21*, 613-619.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Andersson, G., Kaldo-Sandström, V., Ström, L., & Strömgren, T. (2003). Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. *Journal of Psychosomatic Research, 55*, 259-262.

Andrews, G., Slade, T., Peters, L. (1999). Classification in psychiatry: ICD-10 versus DSM-IV. *British Journal of Psychiatry*, *174*, 3-5.

Andrews, G., Sunderland, M., & Kemp, A. (2010). Consistency of diagnostic thresholds in DSM-V. *Australian and New Zealand Journal of Psychiatry, 44*, 309-313.

Angle, H. V., Ellinwood, E. H., & Carroll, F. (1978). *Computer interview problem assessment of psychiatric patients.* Paper presented at The Second Annual Symposium on Computer Application in Medical Care.

Ansseau, M., Fischler, B., Dierick, M., Albert, A., Leyman, S., & Mignon, A. (2008). Socioeconomic correlates of generalized anxiety disorder and major depression in primary care: The GADIS II study. *Depression & Anxiety, 25*, 506-513.

Antony, M. M., Coons, M. J., McCabe, R. E., Ashbaugh, A., & Swinson, R. P. (2006). Psychometric properties of the social phobia inventory: Further evaluation. *Behaviour Research and Therapy, 44*, 1177-1185.

Arksey, H., & Knight, P. (1999). *Interviewing for social scientists*. London: Sage.

Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20*(2), 112-119.

Askwall, S. (1985). Computer supported reading vs. reading text on paper: A comparison of two reading situations. *International Journal of Man-Machine Studies*, *22*, 425-439.

Attia, J. (2003). Moving beyond sensitivity and specificity: Using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber, 26*, 111-113.

Austin, D. W., Carlbring, P., Richards, J. C., & Andersson, G. (2006). Internet administration of three commonly used questionnaires in panic research: Equivalence to paper administration in Australian and Swedish samples of people with panic disorder. *International Journal of Testing, 6*, 25 – 39.

Australian Institute of Health and Welfare. (2011). *Australia's health 2008* (Report No. AUS 99). Canberra: AIHW. Retrieved from http://www.aihw.gov.au/publication-detail/?id=6442468102>

Australian Psychological Society. (2010). *Evidence-based psychological interventions in the treatment of mental disorders: A literature review*. Retrieved from Australian Psychological Society website: www.psychology.org.au.

Australian Psychological Society. (2013). *APS 2012-2013 Schedule of recommended fees and item numbers for psychological services*. Retrieved May 2013 from http://www.psychology.org.au/community/fees_rebates/

Ayearst, L. E., & Bagby, M. R. (2010). Evaluating the psychometric properties of psychological measures. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders*. New York: The Guilford Press.

Ayers, J. W., Althouse, B. M., Allem, J.-P., Rosenquist, J. N., & Ford, D. E. (2013). Seasonality in seeking: Mental health information on Google. *American Journal of Preventive Medicine*, *44*, 520-525.

Aziz, M. A., & Kenford, S. (2004). Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. *Journal of Psychiatric Practice, 10*, 307-313.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The Alcohol Use Disorders Identification Test: Guidelines for use in primary care*. Geneva: World Health Organization, Department of Mental Health and Substance Dependence.

Bachofen, M., Nakagawa, A., Marks, I. M., Park, J. M., Greist, J. H., Baer, L., … Dottl, S. L. (1999). Home self-assessment and self-treatment of obsessive-compulsive disorder using a manual and a computer-conducted telephone interview: Replication of a UK-US study. *Journal of Clinical Psychiatry, 60*, 545-549.

Baillie, A. J., & Rapee, R. M. (2004). Predicting who benefits from psychoeducation and self help for panic attacks. *Behaviour Research and Therapy 42*, 513-527.

Baker, L., Wagner, T. H., Singer, S., & Bundorf, M. K. (2003). Use of the internet and e-mail for health care information. *The Journal of the American Medical Association, 289*, 2400-2406.

Bambling, M., King, R., Reid., W., & Wegner, K. (2008). Online counselling: The experience of counsellors providing synchronous single-session counselling to young people. *Counselling and Psychotherapy Research*, *8*, 110-116.

Barak, A., & Buchanan, T. (2004). Internet-based psychological testing and assessment. In R. Kraus, G. Stricker, & J. Zack (Eds.), *Online counseling: A handbook for mental health professionals* (pp. 217-239). San Diego, CA: Elsevier Academic Press.

Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services, 19* (2/3), 65-89.

Barak, A., & Hen, L. (2008). Exposure in cyberspace as means of enhancing psychological assessment. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications.* (pp. 129-162). Cambridge, UK: Cambridge University Press.

Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services, 26*(2-4), 109-160.

Barak, A., Klein, B., & Proudfoot, J. (2009). Defining Internet-supported therapeutic interventions. *Annals of Behavioral Medicine, 38*, 4-17.

Bargas-Avila, J. A., & Hornbaek, K. (2011). *Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience*. Paper presented at the ACM Conference on Human Factors in Computing Systems.

Bargh, J. A., McKenna, K. Y., & Fitzsimons, G. M. (2002). Can you see the real me? Activation and expression of the “true self” on the Internet. *Journal of Social Issues, 58*, 33.

Baronet, A. M. (1999). Factors associated with caregiver burden in mental illness: A critical review of the research literature. *Clinical Psychological Review, 19*, 819-841.

Barron, M. R., Daniels, J. L., & O'Toole, W. M. (1987). The effect of computer-conducted versus counselor-conducted initial intake interviews on client expectancy. *Computers in Human Behavior, 3*, 21-28.

Bartram, D. (1997). Distance assessment: Psychological assessment through the Internet. *Selection Development Review, 13*, 10-14.

Bartram, D. (2002). Testing on the internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances*. Chichester: John Wiley & Sons Ltd.

Bartram, D. (2006). The internationalization of testing and new models of test delivery on the Internet. *International Journal of Testing, 6*, 121-131.

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2*, 297-307.

Bayliss, M. S., Dewey, J. E., Dunlap, I., Batenhorst, A. S., Cady, R., Diamond, M. L., & Sheftell, F. (2003). A study of the feasibility of Internet administration of a computerized health survey: The headache impact test (HIT). *Quality of Life Research, 12*, 953.

Baxter, J., & Eyles, J. (1997). Evaluating qualitative research in social geography: Establishing 'rigour' in interview analysis. *Transaction of the Institute of British Geographers, 22*, 505-525.

Beckers, J. J., Wicherts, J. M., & Schmidt, H. G. (2007). Computer anxiety: "Trait" or "state"? *Computers in Human Behavior, 23*, 2851-2862.

Bedard-Gilligan, M., Jaeger, J., Echiverri-Cohen, A., & Zoellner, L. A. (2012). Individual differences in trauma disclosure. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 716-723.

Begg, S. J., Vos, T., Barker, B., Stanley, L., & Lopez, A. D. (2008). Burden of disease and injury in Australia in the new millennium: Measuring health loss from diseases, injuries and risk factors. *Medical Journal of Australia, 188*, 36-40.

Bennett, K., Reynolds, J., Christensen, H., & Griffiths, K. M. (2010). E-hub: An online self-help mental health service in the community. *Medical Journal of Australia*, *192*(11), S48-S52.

Bensley, R. J., & Lewis, J. B. (2002). Analysis of internet-based assessments. *Health Promotion Practice 3*, 463-476.

Ben-Ze'ev, A. (2003). Privacy, emotional closeness, and openness in cyberspace. *Computers in Human Behavior, 19*, 451-467.

Berger, M., Wagner, T. H., & Baker, L. C. (2005). Internet use and stigmatized illness. *Social Science & Medicine, 61*, 1821-1827.

Bernadt, M. W., Daniels, O. J., Blizard, R. A., & Murray, R. M. (1989). Can a computer reliably elicit an alcohol history? *British Journal of Addiction, 84*, 405-411.

Berry, M., & Linoff, G. (2000). Mastering data mining. *The art and science of customer relationship management*. New York: John Wiley & Sons.

Beutler, L. E., Alomohamed, S., Moleiro, C., & Romanelli, R. (2002). Systemic treatment selection and prescriptive therapy. In F. W. Kaslow (Ed.), *Comprehensive handbook of sychotherapy: Integrative/eclectic* (Vol. 4, pp. 255-271). Hoboken: John Wiley & Sons Inc.

Bhugra, D., Easter, A., Mallaris, Y., & Gupta, S. (2011). Clinical decision making in psychiatry by psychiatrists. *Acta Psychiatrica Scandinavica, 124*, 403-411.

Bloom, D. E., Cafiero, E. T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L. R., Fathima, S., … Weinstein, C. (2011). *The global economic burden of non-communicable diseases*. Geneva: World Economic Forum.

Blouin, A. G., Perez, E. L., & Blouin, J. H. (1988). Computerized administration of the diagnostic interview schedule. *Psychiatry Research, 23*, 335-344.

Borzekowski, D. L., Leith, J., Medoff, D. R., Potts, W., Dixon, L. B., … Himelhoch, S. (2009). Use of the internet and other media for health information among clinic outpatients with serious mental illness. *Psychiatric Services*, *60*, 1265-1268.

Bower, P. (2005). Stepped care in psychological therapies: Access, effectiveness and efficiency. *The British Journal of Psychiatry, 186*, 11-17.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*, 281-291.

Boyd, J. H., Burke, J. D., Gruenberg, E., Holzer, C. E., Rae, D. S., George, L. K., Karno, M. … Nestadt, G. (1984). The exclusion criteria of DSM-III: A study of co-occurrence of hierarchy-free syndrome. *Archives of General Psychiatry*, *41*, 983–989.

Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A., & Schulberg, H. C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry, 43*, 435-440.

Brosnan, M., & Davidson, M. (1994). Computerphobia: Is it a particularly female phenomenon? *The Psychologist, 7*(2), 73-78.

Brown, T. A., Di Nardo, P. A., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV)*. San Antonio, TX: Psychological Corporation.

Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology, 110*, 49-58.

Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., … Keller, M. B. (2005). Influence of psychiatric co-morbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: A 12-Year prospective study. *American Journal of Psychiatry, 162*, 1179-1187.

Brugha, T. S. (1999). A difference that matters: Comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychological Medicine, 29*, 1013-1020.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice, 33*, 148-154.

Bufka, L. F., & Camp, N. (2010). Brief measures for screening and measuring mental health outcomes. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 62-94). New York: The Guilford Press.

Bunevicius, A., Peceliuniene, J., Mickuviene, N., Valius, L., & Bunevicius, R. (2007). Screening for depression and anxiety disorders in primary care patients. *Depression and Anxiety, 24*, 455-460.

Burns, J. M., Davenport, T. A., Durkin, L. A., Luscombe, G. M., & Hickie, I. B. (2010). The internet as a setting for mental health service utilisation by young people. *The Medical Journal of Australia, 192*(11), S22-S26.

Butcher, J. N. (2003). Computerized psychological assessment. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (pp. 141-163). New York: John Wiley & Sons.

Butcher, J. N. (2004). Computers in clinical assessment: Historical developments, present status, and future challenges. *Journal of Clinical Psychology, 60*, 331-345.

Butcher, J. N. (2006). Assessment in clinical psychology: A perspective on the past, present challenges, and future prospects, *13*, 205-209.

Byers, A. P. (1981). Psychological evaluation by means of an on-line computer. *Behavior Research Methods & Instrumentation, 13*, 585-587.

Byrne, P. (1999). Stigma of mental illness: Changing minds, changing behaviour. *British Journal of Psychiatry, 174*, 1-2.

Cairney, J., Velduizen, S., Kurdyak, P., Missiuna, C., Faught, B. E., & Hay, J. (2007). Evaluating the CSAPPA subscales as potential screening instruments for developmental coordination disorder. *Archives of Disease in Childhood, 92*, 987-991.

Carlbring, P., Forslin, P., Ljungstrand, P., Willebrand, M. Strandlund, C., Ekselius, L., & Andersson, G. (2002). Is the internet-administered CIDI-SF equivalent to a human SCID-interview? *Cognitive Behaviour Therapy*, *31*, 183-189.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Ost, L.-G., & Andersson, G. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior, 23*, 1421-1434.

Carr, A. C., Ancill, R. J., Ghosh, A., & Margo, A. (1981). Direct assessment of depression by microcomputer. *Acta Psychiatrica Scandinavica, 64*, 415-422.

Carr, A. C., Ghosh, A., & Ancill, R. J. (1983). Can a computer take a psychiatric history? *Psychological Medicine, 13*, 151-158.

Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice, 33*, 446-453.

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

Chinman, M., Young, A., Schell, T., Hassell, J., & Mintz, J. (2004). Computer-assisted self-assessment in persons with severe mental illness. *Journal of Clinical Psychiatry, 65*, 1343-1351.

Chinman, M., Hassell, J., Magnabosco, J., Nowlin-Finch, N., Marusak, S., & Young, A. S. (2007). The feasibility of computerized patient self-assessment at mental health clinics. *Administration and Policy of Mental Health, 34*, 401-409.

Chiu, T., & Eysenbach, G. (2010). Stages of use: Consideration, initiation, utilization, and outcomes of an internet-mediated intervention. *BMC Medical Informatics and Decision Making, 10*(73). Retrieved from http://www.biomedcentral.com/content/pdf/1472-6947-10-73.pdf.

Christensen, H., & Griffiths, K. (2003). The internet and mental health practice. *Evidence Based Mental Health*, *6*, 66-69.

Christensen, H., & Hickie, I. B. (2010). Using e-health applications to deliver new mental health services. *The Medical Journal of Australia, 192*(11), S53-S56.

Christopherson, K. M. (2007). The positive and negative implications of anonymity in internet social interactions: "On the internet, nobody knows you're a dog". *Computers in Human Behavior, 23*, 3038-3056.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Clarke, D. M., & McKenzie, D. P. (2003). An examination of the efficiency of the 12-item SPHERE questionnaire as a screening instrument for common mental disorders in primary care. *The Australian and New Zealand Journal of Psychiatry, 37*, 236-239.

Clark, L. A., & Watson, D. J. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Abnormal Psychology*, *100*, 316-336.

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

Cohn, J. F., Kruez, T. S., Matthews, I., Ying, Y., Minh Hoai, N., Padilla, M. T., … De la Torre, F. (2009). *Detecting depression from facial actions and vocal prosody.* Paper presented at the Affective Computing and Intelligent Interaction and Workshops, 2009.

Coles, M. E., & Coleman, S. L. (2010). Barriers to treatment seeking for anxiety disorders: Initial data on the role of mental health literacy. *Depression & Anxiety, 27*, 63-71.

Coles, M. E., Cook, L. M., & Blake, T. R. (2007). Assessing obsessive compulsive symptoms and cognitions on the internet: Evidence for the comparability of paper and Internet administration. *Behaviour Research and Therapy, 45*, 2232-2240.

Collins, F. E., & Jones, K. V. (2004). Investigating dissociation online: Validation of a web-based version of the Dissociative Experiences Scale. *Journal of Trauma & Dissociation, 5*, 133-147.

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Foa, E., & Wesler, R. H. (2000). Psychometric properties of the social phobia inventory (SPIN). *British Journal of Psychiatry, 176*, 379-386.

Corrigan, P. W., & Watson, A. C. (2002). Understanding the impact of stigma on people with mental illness. *World Psychiatry*, *1*, 16–20.

Cooper, *Z.,* Cooper*,* P. J., & Fairburn, C. G. (1989). The validity of the eating disorder examination and its subscales. *British Journal of Psychiatry*, *154*, 807-812*.*

Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. Y. (2009). Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS One, 4*(7), e6353.

Costafreda, S. G., Fu, C. H. Y., Picchioni, M., Toulopoulou, T., McDonald, C., Kravariti, E., … McGuire, P. K. (2011). Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry, 11*(18). doi: 10.1186/1471-244X-11-18

Costello, A., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(7), 1-9.

Cotton, M.-A., Ball, C., & Robinson, P. (2003). Four simple questions can help screen for eating disorders. *Journal of General Internal Medicine, 18*(1), 53-56.

Coyne, I., Bartram, D., & Smith-Lee Chong, P. (2003). Survey of good practice issues: Internet delivered personality testing in the UK. *Selection Development Review, 19*, 3-6.

Chronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Craemer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the Impact of Event Scale - Revised. *Behaviour Research and Therapy, 41*, 1489-1496.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour and Information Technology*, *6*(1), 3-13.

Crisp, A. H., Gelder, M. G., Rix, S., Meltzer, H. I., Rowlands, O. J. (2000). Stigmatisation of people with mental illnesses. *British Journal of Psychiatry*, *177*, 4-7.

Cuijpers, P., Boluijt, P., & Van Straten, A. (2008). Screening of depression in adolescents through the Internet. *European Child & Adolescent Psychiatry, 17*(1), 32-38.

Cuijpers, P., Li, J., Hofmann, S. G., & Andersson, G. (2010). Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: A meta-analysis. *Clinical Psychological Review, 30*, 768-778.

Cuijpers, P., Smit, F., Hollon, S. D., & Anderson, G. (2010a). Continuous and dichotomous outcomes in studies of psychotherapy for adult depression: A meta-analytic comparison. *Journal of Affective Disorders, 126*, 349-357.

Cunningham, J. A., Humphreys, K., Kypri, K., & van Mierlo, T. (2006). Formative evaluation and three-month follow-up of an online personalized assessment feedback intervention for problem drinkers. *Journal of Medical Internet Research, 8*(2), e5.

Cunningham, J. A., & Van Mierlo, T. (2009). Methodological issues in the evaluation of Internet-based interventions for problem drinking. *Drug & Alcohol Review, 28*(1), 12-17.

Dalal, P. K., & Sivakumar, T. (2009). Moving towards ICD-11 and DSM-V: Concept and evolution of psychiatric classification. *Indian Journal of Psychiatry*, *54*, 310-319.

Daly, A. E. (2005). *Bridging the digital divide: The role of community online access centres in indigenous communities*. Discussion Paper No. 273/2005. Centre for Aboriginal Economic Policy Research.

Danaher, B. G., Mckay, G. H., & Seeley, J. R. (2005). The information architecture of behavior change websites. *Journal of Medical Internet Research, 7*(2), e12.

Danaher, B. G., & Seeley, J. R. (2009). Methodological issues in research on web-based behavioural interventions. *Annual Review of Behaviorual Medicine, 38,* 28-39.

Dart, J. (2008). The internet as a source of health information in three disparate communities. *Australian Health Review, 32*, 559-569.

Davidson, L., Chinman, M., Kloos, R. W., Stayner, D., & Tebes, J. K. (1999). Peer support among individuals with severe mental illness: A review of the evidence. *Clinical Psychology: Science and Practice, 6*, 165-187.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behaviour Research Methods, Instruments, & Computers, 31*, 572-577.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review, 112*(1), 129-158.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics, 35*, 1297-1326.

Di Nardo, P. A., Moras, K., Barlow, D. H., Rapee, R. M., & Brown, T. A. (1993). Reliability of DSM-III-R anxiety disorder categories: Using the anxiety disorders interview schedule-revised (ADIS-R). *Archives of General Psychiatry, 50*, 251-256.

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2009). A brief web screening questionnaire for common mental disorders: Development and validation. *Journal of Medical Internet Research, 11*(3), e19.

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2010). Brief self-rated screening for depression on the internet. *Journal of Affective Disorders, 122*, 253-259.

Donker, T., van Straten, A., Marks, I., & Cuijpers, P. (2011). Quick and easy self-rating of Generalized Anxiety Disorder: Validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. *Psychiatry Research, 188*(1), 58-64.

Dozois, D. J., & Dobson, K. S. (2010). Depression. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment for psychological disorders* (2nd ed., pp. 344-389). New York: Guilford Press.

Dozois, D. J., Dobson, K. S., & Ahnberg, J. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10*, 83-89.

Draine, J., Salzer, M. S., Culhane, D. P., & Hadley, T. R. (2002). Role of social disadvantage in crime, joblessness, and homelessness among persons with serious mental illness. *Psychiatric Services, 53*, 565-573.

Ebert, A., & Bär, K.-J. (2010). Emil Kraepelin: A pioneer of scientific understanding of psychiatry and psychopharmacology. *Indian Journal of Psychiatry, 52*, 191-192.

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology, 38*, 215-229.

Emmelkamp, P. M. G. (2005). Technological innovations in clinical assessment and psychotherapy. *Psychotherapy and Psychosomatics, 74*, 336-343.

Endicott, J. (2001). Good diagnoses require good diagnosticians: Collecting and integrating the data. *American Journal of Medical Genetics, 105*(1), 48-49.

Erdman, H. P., Klein, M. H., Greist, J. H., Skare, S. S., Husted, J. J., Robins, L. N., … Phillip, J. (1992). A comparison of two computer-administered versions of the NIMH diagnostic interview schedule. *Journal of Psychiatric Research, 26*(1), 85-95.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.

Erlanger, D. M., Kaushik, T., Broshek, D., Freeman, J., Feldman, D., & Festa, J. (2002). Development and validation of a web-based screening tool for monitoring cognitive status. *Journal of Head Trauma Rehabilitation, 17*, 458-476.

Etter, J.-F. (2005). A self-administered questionnaire to measure cigarette withdrawal symptoms: The Cigarette Withdrawal Scale. *Nicotine & Tobacco Research, 7*, 47-57.

Evans, M., Robling, M., Maggs, R., Houston, H., Kinnersley, P., & Wilkinson, C. (2002). It doesn't cost anything just to ask, does it? The ethics of questionnaire-based research. *Journal of Medical Ethics, 28*, 41-44.

Eyde, L. D., & Kowal, D. M. (1987). Computerised test interpretation services: Ethical and professional concerns regarding U.S. producers and users. *Applied Psychology: An International Review, 36*, 401-417.

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research, 7*(1), e11.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal (Clinical Research Edition), 324*, 573-577.

Fairburn C. G., & Cooper, Z. (1993). The Eating Disorder Examination (12th edition). In: C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment and treatment* (pp. 317-360). New York: Guilford Press, 1993.

Faja, S., & Likcani, A. (2006). E-Health: An exploratory study of trust building elements in behavioural health web sites. *Journal of Information Science and Technology, 3*(1), 9-21.

Fallaw, S. S., Kantrowitz, T. M., & Dawson, C. R. (2012). *2012 Global Assessment Trends Report*. Retrieved from http://www.shl.com/assets/GATR_2012_US.pdf.

Fallows, D. (2005). *Search engine users*. Washington, DC: Pew Internet and American Life Project.

Farvolden, P., McBride, C., Bagby, R. M., & Ravitz, P. (2002). A web-based screening instrument for anxiety and depression disorders in primary care. *Journal of Medical Internet Research, 5*(23), e23.

Faye, G. (2005). Stigma: Barrier to mental health care among ethnic minorities. *Issues in Mental Health Nursing, 26*, 979-999.

Feighner, J. P., Robins, E., Guze, S. B., Woodrmff, E. A., Winokur, G., & Mum, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26*, 57-63.

Fennell, D., & Liberato, A. S. (2007). Learning to live with OCD: Labeling, the self, and stigma. *Deviant Behavior, 28*, 305-331.

Ferrando, P. J., & Lorenzo-Seva, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and internet-administered questionnaires. *Psychological Methods, 10*, 193-205.

Ferreira-Lay, P., & Miller, S. (2008). The quality of internet information on depression for lay people. *Psychiatric Bulletin, 32*, 170-173.

Ferriter, M. (1993). Computer aided interviewing and the psychiatric social history. *Social Work and Social Sciences Review, 4*, 255-263.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.

Finlay, L. (2009). Debating phenomenological research methods. *Phenomenology & Practice, 3*(1), 6-25.

Finger, M. S., & Ones, D. S. (1999). Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. *Psychological Assessment, 11*, 58-66.

First, M. B., & Westen, D. (2007). Classification for clinical practice: How to make ICD and DSM better able to serve clinicians. *International Review of Psychiatry, 19*, 473-481.

First, M. B. (2010). Clinical utility in the revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM). *Professional Psychology: Research and Practice, 41*, 465-473.

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care in Medicine, 29*, 1043-1051.

Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., & Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT) *Quality of Life Research, 14*, 2277-2291.

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. *Journal of Traumatic Stress, 6*, 459-473.

Fogel, J., & Ford, D. E. (2005). Stigma beliefs of Asian Americans with depression in an internet sample. *The Canadian Journal of Psychiatry*, *50*, 470-478.

Fortson, B. L., Scotti, J. R., Del Ben, K. S., & Chen, Y.-C. (2006). Reliability and validity of an internet traumatic stress survey with a college student sample. *Journal of Traumatic Stress, 19*, 709-720.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*, 248-260.

Fowler, R. D., & Miller, M. L. (1969). Computer interpretation of the MMPI: Its use in clinical practice. *Archives of General Psychiatry, 21*, 502-508.

Fox, S., Rainie, L., Horrigan, J., Lenhart, J., Lenhart, A., Spooner, T., … Carter, C. (2000). The online health care revolution: How the web helps Americans take better care of themselves. *The Pew Internet & American Life Project*. Washington, DC: The Pew Internet & American Life Project.

Frances, A. (2009). Whither DSM-V. *The British Journal of Psychiatry, 195*, 391-392.

Fray, P. J., Robbins, T. W., & Sahakian, B. J. (1996). Neuropsychological applications of CANTAB. *International Journal of Geriatric Psychiatry*, *11*, 329–336.

Free, C., Phillips, G., Galli, L., Watson, L., Felix, L., Edwards, P., … Haines, A. (2013). The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. *PLoS Med, 10*(1), e1001362.

Fries, J. F., Bruce, B., & Cella, D. (2005). The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. *Clinical and Experimental Rheumatology, 23*(Suppliment 39), 53-57.

Gainsbury, S. M., & Blaszcynski, A. (2011). A systematic review of Internet-based therapy for the treatment of addictions. *Clinical Psychology Review, 31*, 480-498.

Galaif, E. R., Sussman, S., Newcomb, M. D., & Locke, T. F. (2007). Suicidality, depression, and alcohol use among adolescents: A review of empirical findings. *International Journal of Adolescent Medicine and Health, 19*, 27-35.

Galdas, P. M., Cheater, F., & Marshall, P. (2005). Men and health help-seeking behaviour: Literature review. *Journal of Advanced Nursing, 49*, 616-623.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67-89.

Garb, H. N., Lillienfield, S., & Fowler, K. A. (2008). Psychological assessment and clinical judgment. In J. E. Maddux & B. A. Winstead (Eds.), *Psychopathology: Foundations for a contemporary understanding* (2nd ed.). New York: Routledge.

García-Lizana, F., & Muñoz-Mayorga, I. (2010). What about telepsychiatry? A systematic review. *Journal of Clinical Psychiatry, 12*, 2-6.

Gardner, W., Shear, K., Kelleher, K. J., Pajer, K. A., Mammen, O., Buysse, D., & Frank, E. (2004). Computerized adaptive measurement of depression: A simulation study. *BMC Psychiatry, 6*, 4-13.

Gaynes, B. N., DeVeaugh-Geiss, J., Weir, S., Gu, H., MacPherson, C., Schulberg, H. C., … Rubinow, M. (2010). Feasibility and diagnostic validity of the M-3 checklist: A brief, self-rated screen for depressive, bipolar, anxiety, and post-traumatic stress disorders in primary care. *Annals of Family Medicine, 8*, 160-169.

Gega, L. (2004). Computer-aided CBT self-help for anxiety and depressive disorders: Experience of a London clinic and future directions. *Journal of Clinical Psychology, 60*, 147-157.

Geisser, M. E., Roth, R. S., & Robinson, M. E. (1997). Assessing depression among persons with chronic pain using the Center for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: A comparative analysis. *The Clinical Journal of Pain, 13*, 163-170.

George, C. E., Lankford, J. S., & Wilson, S. E. (1992). The effects of computerized versus paper-and pencil administration on measures of negative affect. *Computers in Human Behavior, 8*, 203-209.

Geraghty, A. W., Wood, A. M., & Hyland, M. E. (2010). Attrition from self-directed interventions: Investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. *Social Science and Medicine*, *71*(1), 30-7.

Gibson, C. (1998). Semi-structured and unstructured interviewing: A comparison of methodologies in research with patients following discharge from an acute psychiatric hospital. *Journal of Psychiatric and Mental Health Nursing, 5*, 469-477.

Giorgi. (1997). The theory, practice, and evaluation of the phenomenological method as a qualitative research procedure. *Journal of Phenomenological Psychology, 28*, 235-260.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*, 597-607.

Goldberg, L., Lide, B., Lowry, S., Massett, H. A., O'Connell, T., Preece, J., … Shneiderman, B. (2011). Usability and accessibility in consumer health informatics: Current trends and future challenges. *American Journal of Preventive Medicine, 40*(5, Supp. 2), 187-197.

Goldner, M. (2006). How health status impacts the types of information consumers seek online. *Information, Communication & Society, 9*, 693-713.

Gonzalez, G. M., & Shriver, C. (2004). A bilingual computerized voice-interactive system for screening depression symptoms. *Journal of Technology in Human Services, 22*(4), 1-20.

Goodwin, R. D., Lieb, R., Hoefler, M., Pfister, H., Bittner, A., Beesdo, K., & Wittchen, H. U. (2004). Panic attack as a risk factor for severe psychopathology. *American Journal of Psychiatry, 161*, 2207-2214.

Gould, R. A., & Clum, G. A. (1993). A meta-analysis of self-help treatment approaches. *Clinical Psychology Review, 3*, 169-186.

Grady, B., Myers, K., Nelson, E. L. (2009). Telemental Health Standards and Guidelines Working Group: Evidence based Practice for Telemental Health. American Telemedicine Association; Evidence Based Practice for Telemental Health.

Greist, J. H., Klein, M. H., Erdman, H. P., & Jefferson, J. W. (1982). Clinical computer applications in mental health. *Proceedings American Medical Informatics Association Symposium, 2*, 356-365.

Grieve, R., & de Groot, H. T. (2011). Does online psychological test administration facilitate faking? *Computers in Human Behavior*, *27*, 2387-2391.

Grieve, R., & Elliott, J. (2013). Cyberfaking: I can, so I will? Intentions to fake in online psychological testing. *CyberPsychology, Behavior, and Social Networking*, *16*, 364-369.

Griffis, S. E., Goldsby, T. J., & Cooper, M. (2003). Web-based and mail surveys: A comparison of response, data, and cost. *Journal of Business Logistics, 24*, 237-258.

Griffiths, K. M., Farrer, L., & Christensen, H. (2010). The efficacy of internet interventions for depression and anxiety disorders: A review of randomised controlled trials. *Medical Journal of Australia*, *192*(11), S4-S11

Griffiths, F., Lindenmeyer, A., Powell, J., Lowe, P., & Thorogood, M. (2006). Why are health care interventions delivered over the internet? A systematic review of the published literature. *Journal of Medical Internet Research, 8*(2), e10.

Grisham, J. R., Brown, T. A., & Campbell, L. A. (2004). The Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV). In M. J. Hilsenroth, D. L. Segal & M. Hersen (Eds.), *Comprehensive handbook of psychological assessment, personality assessment* (Vol. 2). Hoboken: John Wiley & Sons.

Groth-Marnat, G. (2003). *Handbook for psychological assessment* (4th ed.). Hoboken: John Wiley & Sons, Inc.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Guarda, A., & Redgrave, G. W. (2004). Eating disorders: Detection, assessment, and treatment in primary care. *Advanced Studies in Medicine, 4*, 468-475.

Hanna, R. C., Weinberg, B., Dant, R. P., & Berger, P. D. (2005). Do internet-based surveys increase personal self-disclosure? *Journal of Database Marketing & Customer Strategy Management, 12*, 342-356.

Harwood, T. M., Pratt, D., Beutler, L. E., Bongar, B. M., Lenore, S., & Forrester, B. T. (2011). Technology, telehealth, treatment enhancement, and selection. *Professional Psychology: Research and Practice, 42*, 448-454.

Hassenzahl, M. (2008). *User experience (UX): Towards an experiential perspective on product quality.* Paper presented at the IMH 2008 20th International Conference of the Association Francophone d'Interaction Homme-Machine.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology, 25*, 91-97.

Hawton, K., Salkovskis, P. M., Kirk, J., & Clarke, D. M. (2010). *Cognitive behaviour therapy for psychiatric problems: A practical guide.* Oxford: Oxford University Press.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experimental avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting & Clinical Psychology, 64*, 1152-1168.

Hayne, Y. (2003). Experiencing psychiatric diagnosis: Client perspectives on being named mentally ill. *Journal of Psychiatric and Mental Health Nursing, 10*, 722-729.

Hayne, S., & Rice, R. (1997). Attribution accuracy when using anonymity in group support systems. *International Journal of Human Computer Studies*, *47*, 429-50.

Hemelrijk, E., van Ballegooijen, W., Donker, T., van Straten, A., & Kerkhof, A. (2012). Internet-based screening for suicidal ideation in common mental disorders. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 33*, 215-221.

Henderson, A. S., & Jorm, A. F. (1990). Do mental health surveys disturb? *Psychological Medicine, 20*, 721-724.

Herman, S., & Koran, L. M. (1998). In vivo measurement of obsessive-compulsive disorder symptoms using palmtop computers. *Computers in Human Behavior, 14*, 449-462.

Herrero, J., & Meneses, J. (2006). Short web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CES-D) Scales: A comparison to pencil and paper responses among Internet users. *Computers in Human Behavior, 22*, 830-846.

Hilsenroth, M. J., Segal, D. L., & Hersen, M. (2004). *Comprehensive handbook of psychological assessment: Personality assessment*. Hoboken: John Wiley & Sons.

Hilton, M. F., Scuffham, P. A., Vecchio, N., & Whiteford, H. A. (2010). Using the interaction of mental health symptoms and treatment status to estimate lost employee productivity. *Australian & New Zealand Journal of Psychiatry, 44*, 151-161.

Hofmann, S. G., & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69*, 621-632.

Hollon, S. D. (2011). Cognitive and behavior therapy in the treatment and prevention of depression. *Depression & Anxiety, 28*, 263-266.

Holländare, F., Andersson, G., & Engström, I. (2010). A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *Journal of Medical Internet Research, 12*(5), e49.

Hosmer, D. W., & Lemeshow, S. (2000), *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.

Houck, P. R., Spiegel, D. A., Shear, M. K., & Rucci, P. (2002). Reliability of the self-report version of the panic disorder severity scale. *Depression and Anxiety, 15*, 183-185.

Houston, T. K., Cooper, L. A., Vu, H. T., Kahn, J., Toser, J., & Ford, D. E. (2001). Screening the public for depression through the Internet. *Psychiatric Services, 52*, 362-367.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277-1288.

Hsu, L. M. (2002). Diagnostic validity statistics and the MCMI-III. *Psychological Assessment*, *14*, 410-422.

Hungsley, J. (2003). Cost effectiveness and medical cost-offset considerations in psychological service provision. *Canadian Psychology*, *44*, 61-73.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29-51.

Hunsley, J., & Mash, E. J. (2010). The role of assessment in evidence-based practice. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 3-22). New York: The Guilford Press.

Hyler, S. E., Gangure, D. P., & Batchelder, S. T. (2005). Can telepsychiatry replace in-person psychiatric assessments? A review and meta-analysis of comparison studies. *Central Nervous System Spectrum*, *10*, 403-413.

International Testing Commission. (2005). *International guidelines on computer-based and internet delivered testing*. Retrieved from http://www.intestcom.org/Downloads/ITC%20Guidelines%20on%20Computer%20-%20version%202005%20approved.pdf.

Jacomb, P. A., Jorm, A. F., Rodgers, J., Korten, A. E., Henderson, A. S., & Christensen, H. (1999). Emotional response of participants to a mental health survey. *Social Psychiatry and Psychiatric Epidemiology, 34*, 80-84.

Jardi, R., Pelta, J., Maron, M., Thomas, P., Delion, P., Codaccioni, X., & Goudemand, M. (2006). Predictive validation study of the Edinburgh Postnatal Depression Scale in the first week after delivery and risk analysis for postnatal depression. *Journal of Affective Disorders, 93*, 169-176.

Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology, 76*, 711-722.

Jensen-Doss, A., & Hawley, K. M. (2011). Understanding clinicians' diagnostic practices: attitudes toward the utility of diagnosis and standardized diagnostic tools. *Administration and Policy of Mental Health, 38*, 476-485.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*, 112-133.

Joiner, T. E., & Pettit, J. W. (2008). Adult clinical assessment and diagnosis research: Current status and future research. In M. Roberts & S. S. Ilardi (Eds.), *Handbook of Research Method in Clinical Psychology*. Oxford: John Wiley & Sons, Ltd.

Joinson, A. N. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*, 433-438.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*, 177-192.

Joinson, A. N., & Paine-Schofield, C. B. (2007). Self-Disclosure, privacy and the internet. In A. N. Joinson, K. Y. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford Handbook of Internet Psychology* (pp. 237-252). Oxford: Oxford University Press.

Joling, K. J., van Marwijk, H. W., Piek, E., van der Horst, H. E., Penninx, B. W., Verhaak, P., & van Hout, H. P. J. (2011). Do GPs' medical records demonstrate a good recognition of depression? A new perspective on case extraction. *Journal of Affective Disorders, 133*, 522-527.

Jones, K. D. (2010). The unstructured clinical interview. *Journal of Counseling & Development, 88*, 220-226.

Jones, S. H., Thornicroft, G., Coffey, M., Dunn, G. (1995). A brief mental health outcome scale-reliability and validity of the Global Assessment of Functioning (GAF). *British Journal of Psychiatry, 166*, 654-659.

Jorm, A. F., Henderson, A. S., Scott, R., Mackinnon, A., Korten, A. E., & Christensen, H. (1994). Do mental health surveys disturb? Further evidence. *Psychological Medicine, 24*, 233-237.

Jorm, A. F., & Reavley, N. J. (2012). Changes in psychological distress in Australian adults between 1995 and 2011. *Australian & New Zealand Journal of Psychiatry, 46*, 352-356.

Joshi, J., Dhall, A., Goecke, R., Breakspear, M., & Parker, G. (2012). *Neural-net classification for spatio-temporal descriptor based depression analysis.* Paper presented at the International Conference on Pattern Recognition ICPR2012, Tsukuba, Japan, 11-15 Nov 2012.

Jowett, A., Peel, E., & Shaw, R. (2011). Online interviewing in psychology: Reflections on the process. *Qualitative Research in Psychology*, *8*, 354-369.

Kang, Y. J., Ku, J., Han, K., Kim, S., Yu, T. W., Lee, J. H., & Park, C. I. (2008). Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: Preliminary study. *CyberPsychology & Behavior, 11*, 329-339.

Katzelnick, D. J., Kobak, K. A., Greist, J. H., Jefferson, J. W., Mantle, J. M., & Serlin, R. C. (1995). Sertraline for social phobia: A double-blind, placebo-controlled crossover study. *American Journal of Psychiatry, 152*, 1368-1371.

Katzelnick, D. J., Kobak, K. A., Greist, J. H., Jefferson, J. W., & Henk, H. J. (1997). Effect of primary care treatment of depression on service use by patients with high medical expenditures. *Psychiatric Services*, *48*, 59-64.

Kauer, S., Reid, S., Sanci, L., & Patton, G. C. (2009). Investigating the utility of mobile phones for collecting data about adolescent alcohol use and related mood, stress and coping behaviours: Lessons and recommendations. *Drug Alcohol Review, 28*, 25-30.

Kays, K., Gathercoal, K., & Buhrow, W. (2011). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior, 28*, 251-256.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*, 1123-1134.

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*(1), 4-12.

Kendler, K. (2009). An historical framework for psychiatric nosology. *Psychological Medicine, 39*, 1935-1941.

Kenardy, J., Fried, L., Kraemer, H. C., & Taylor, C. B. (1992). Psychological precursors of panic attacks. *British Journal of Psychiatry, 160*, 668-673.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*, 184-189.

Kessler, R. C., Berglund, P. A., Bruce, M. L., Koch, R., Laska, E. M., Leaf, P. J., … Wang, P. S. (2001). The prevalence and correlates of untreated serious mental illness. *Health Services Research, 36*, 987-1007.

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E.,… Zaslavsky, A. M. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine, 352*, 2515-2523.

Kessler, R. C., & Mroczek, D. (1994). *Final version of our non-specific psychological distress scale*. School of Survey Research Center of the Institute for Social Research, University of Michigan.

Kessler, R. C., & Ustün, T. B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research, 13*, 93-121.

Khan, A., Leventhal, R. M., Khan, S., & Brown, W. A. (2002). Suicide risk in patients with anxiety disorders: A meta-analysis of the FDA database. *Journal of Affective Disorders, 68*, 183-190.

Khan, S., & VanWynsberghe, R. (2008). Cultivating the under-mined: Cross-case analysis as knowledge mobilization. *Forum: Qualitative Social Research, 9*(1), 34.

Khazaal, Y., Chatton, A., Cochand, S., Hoch, A., Khankarli, M., Khan, R., & Zullino, D. (2008). Internet use by patients with psychiatric disorders in search for general and medical informations. *Psychiatric Quarterly, 79*, 301-309.

Kishore, R. V., Samar, R., Reddy, J., Chandrasekhar, C. R., & Thennarasu, K. (2004). Clinical characteristics and treatment response in poor and good insight obsessive-compulsive disorder. *European Psychiatry, 19*, 202-208.

Kim, N. S., & Ahn, W.-K. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology, 131*, 451-476.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine, 5*(2), e45.

Klein, B., Austin, D., Nguyen, D., Smith, A., Donelan, P., Abbott, J., Kyrios, M., … Williams, B. (2012). The first two years of Anxiety Online, an international e-mental health service for the anxiety disorders. In D. Castle, S. Hood, & V. Starcevic (Eds.), *Anxiety disorders:*

Current understandings, novel treatments. Melbourne, VIC: Australian Postgraduate Medicine.

Klein, B., Austin, D. W., & Richards, J. (2001). A brief internet-based treatment for panic disorder. *Behavioural and Cognitive Psychotherapy*, *29*, 113-117.

Klein, B., Meyer, D., Austin, D. W., & Kyrios, M. (2011). Anxiety Online - A virtual clinic: Preliminary outcomes following completion of five fully automated treatment programs for anxiety disorders and symptoms. *Journal of Medical Internet Research, 13*(4), e89.

Klein, B., White, A., Kavanagh, D., Shandley, K., Kay-Lambkin, F., Proudfoot, J., … Young, R. (2010). Content and functionality of alcohol and other drug websites: Results of an online survey. *Journal of Medical Internet Research, 12*(5), e51.

Kobak, K. A., Greist, J. H., Jefferson, J. W., & Katzelnick, D. J. (1996). Computer-administered clinical rating scales: A review. *Psychopharmacology, 127*, 291-301.

Kobak, K. A., Reynolds, W. M., & Greist, J. H. (1993). Development and validation of a computer-administered version of the Hamilton Rating Scale. *Psychological Assessment, 5*, 487-492.

Kobak, K. A., Reynolds, W. M., & Greist, J. H. (1994). Computerized and clinician assessment of depression and anxiety: Respondent evaluation and satisfaction. *Journal of Personality Assessment, 63*, 173.

Kobak, K., Reynolds, W. M., Rosenfeld, R., & Greist, J. H. (1990). Development and validation of a computer administered version of the Hamilton Depression Rating Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 56-63.

Kobak, K. A., Williams, J. B., Jeglic, E., Salvucci, D., & Sharp, I. R. (2008). Face-to-face versus remote administration of the Montgomery-Asberg Depression Rating Scale using videoconference and telephone. *Depression & Anxiety, 25*, 913-919.

Koen, N., & Stein, D. J. (2011). Pharmacotherapy of anxiety disorders: A critical review. *Dialogues in Clinical Neuroscience, 13*, 423-437.

Kontos, E. Z., Bennett, G. G., & Viswanath, K. (2007). Barriers and facilitators to home computer and internet use among urban novice computer users of low socioeconomic position. *Journal of Medical Internet Research, 9*(4), e31.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment

and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*, 1566-1577.

Kraemer, H. C., Noda, A., & O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: Methodological challenges. *Journal of Psychiatric Research, 38*, 17-25.

Kraemer, H. C. (2007). DSM categories and dimensions in clinical and research contexts. *International Journal of Methods in Psychiatric Research, 16*, 8-15.

Krahn, G., & Putnam, M. (2008). Qualitative methods in psychological research. In M. Roberts & S. S. Ilardi (Eds.), *Handbook of Research Methods in Clinical Psychology*. Oxford: Wiley-Blackwell.

Kranzler, H. R., Tennen, H., Babor, T. F., Kadden, R. M., & Rounsaville, B. J. (1995). Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *Drug and Alcohol Dependence, 45*, 93-104.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2*, 111-133.

Krueger, R. F., McGue, M., & Iacono, W. G. (2001). The higher-order structure of common DSM mental disorders: internalization, externalization, and their connections to personality. *Personality and Individual Differences, 30*, 1245-1259.

Kurt, R., Bogner, H. R., Straton, J. B., Tien, A. Y., & Gallo, J. J. (2004). Computer-assisted assessment of depression and function in older primary care patients. *Computer Methods and Programs in Biomedicine, 73*, 165-171.

Kuyken, W., & Tsivrikos, D. (2009). Therapist competence, comorbidity and cognitive-behavioral therapy for depression. *Psychotherapy and Psychosomatics, 78*, 42-48.

Kvale, S. (2007). *Doing interviews*. Thousand Oaks, CA: Sage.

Lahr, D., Beblo, T., & Hartje, W. (2007). Cognitive performance and subjective complaints before and after remission of major depression. *Cognitive Neuropsychiatry, 12*, 25-45.

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain, 8*, 221-223.

Lam-Po-Tang, J., & McKay, D. (2010). Dr Google, MD: A survey of mental health-related internet use in a private practice sample. *Australasian Psychiatry, 18*, 130-133.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lawrence, D. M., Holman, C. D., Jablensky, A. V., & Hobbs, M. S. (2003). Death rate from ischaemic heart disease in Western Australian psychiatric patients, 1980-1998. *British Journal of Psychiatry, 182*, 31-36.

Lea, L. M., & Parker, R. A. (1997). *Designing and conducting survey research: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Leach, L. J., Christensen, H., Griffiths, K., Jorm, A. F., & Mackinnon, A. (2007). Websites as a mode of delivering mental health information: Perceptions from the Australian public. *Social Psychiatry & Psychiatric Epidemiology*, *42*, 167-172.

Leckman, J. F., Denys, D., Simpson, H. B., Mataix-Cols, D., Hollander, E., Saxena, S., … Stein, D. J. (2010). Obsessive–compulsive disorder: A review of the diagnostic criteria and possible subtypes and dimensional specifiers for DSM-V. *Depression & Anxiety, 27*, 507-527.

Lecrubier, Y. (2001). The burden of depression and anxiety in general medicine. *Journal of Clinical Psychiatry, 62*(Supp. 8), 4-9.

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Harnett Sheehan, K., … Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry, 12*, 224-231.

Leibert, T., Archer, J. J., Munson, J., & York, G. (2006). An exploratory study of client perceptions of internet counseling and the therapeutic alliance. *Journal of Mental Health Counseling, 28*, 69-83.

Lemperiere, T. (1995). The importance of classifications in psychiatry. *Encephale*, *5*, 3-7.

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *The American Journal of Psychiatry, 144*, 1184-1188.

Lester, S. (1999). An introduction to phenomenological research. Retrieved from www.sld.demon.co.uk/resmethy.pdf

Levitan, R. D., Blouin, A. G., Navarro, J. R., & Hill, J. (1991). Validity of the computerized DIS for diagnosing psychiatric inpatients. *Canadian Journal of Psychiatry, 36*, 728-731.

Lewis, G. (1994). Assessing psychiatric disorder with a human interviewer or a computer. *Journal of Epidemiological and Community Health, 48*, 207-210.

Lin, C.-C., Bai, Y.-M., Liu, C.-Y., Hsiao, M.-C., Chen, J.-Y., Tsai, S.-J., … Li, Y.-C. (2007). Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. *BMC Psychiatry, 7*(1). doi: 10.1186/1471-244X-7-12

Link, B. (1982). Mental patient status, work and income: An examination of the effects of a psychiatric label. *American Sociological Review*, *47*, 202-15.

Litz, B. T., Williams, L., Wang, J., Bryant, R., & Engel, C. C. (2004). A therapist -assisted internet self-help program for traumatic stress. *Professional Psychology: Research and Practice, 35*, 628–634.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology & Biogeography, 17*, 145-151.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*, 266-274.

Lucas, R. W., Mullin, P. J., Luna, C. B., & McInroy, D. C. (1977). Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: A comparison. *The British Journal of Psychiatry, 131*, 160-167.

Lysaker, P. H., Yanos, P. T., & Roe, D. (2010). Association of stigma, self-esteem, and symptoms with concurrent and prospective assessment of social anxiety in schizophrenia. *Clinical Schizophrenia and Related Psychoses, 4*, 41-48.

Maccoby, E. E., & Maccoby. N. A. (1954). *The interview: A tool of social science*. In G. Lindzey (Ed.), *Handbook of Social Psychology* (Vol. 1, pp. 449-487). Cambridge, MA: Addison-Wesley.

Malgady, R. G., Rogler, L. H., & Tryon, W. W. (1992). Issues of validity in the diagnostic interview schedule. *Journal of Psychiatric Research, 26*, 59-67.

Malhotra, S., Chakrabarti, S., & Shah, R. (2013). Telepsychiatry: Promise, potential, and challenges. *Indian Journal of Psychiatry, 55*, 3.

Mallen, M. J., Jenkins, I. M., Vogel, D. L., & Day, S. X. (2011). Online counselling: An initial examination of the process in a synchronous chat environment. *Counselling & Psychotherapy Research, 11*, 220-227.

Maloof, M. (2003). *Learning when data sets are imbalanced and when costs are unequal and unknown*. ICML-2003 Workshop on Learning from Imbalanced Data Sets (II). Retrieved on May 24, 2007 from http://citeseer.ist.psu.edu/cache/papers/cs/27538/http:zSzzSzwww.cs.georgetown.eduzSz~maloofzSzpubszSzicml03-wids.pdf/maloof03learning.pdf

Markham, D., & Trower, P. (2003). The effects of the psychiatric label 'borderline personality disorder' on nursing staff's perceptions and causal attributions for challenging behaviours. *British Journal of Clinical Psychology, 42*, 243-256.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic patients. *Behaviour Research and Therapy, 17*, 263-267.

Marshall, R. D., Spitzer, R. L., Vaughan, S. C., Vaughan, R., Mellman, L. A., MacKinnon, R. A., & Roose, S. P. (2001). Assessing the subjective experience of being a participant in psychiatric research. *American Journal of Psychiatry, 158*, 319-321.

Matheson, K., & Zanna, M. P. (1988). The impact of computer-mediated communication on self-awareness. *Computers in Human Behavior, 4*, 221-233.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *Public Library of Science, 3*(11), e442.

Mathison, S. (1988). Why triangulate? *Educational Researcher, 17*(2), 13-17.

Mauer, B. (2006). *Morbidity and mortality in people with serious mental illness*. Alexandra, VA: National Association of State Mental Health Program Directors Medical Directors Council.

Mavissakalian, M. (1986). The fear questionnaire: A validity study. *Behaviour Research and Therapy, 24*, 83-85.

Mayes, D. K., Sims, V. K., & Koonce, J. M., (2001). Comprehension and workload differences for VDT and paper-based reading. *International Journal of Industrial Ergonomics*, *28*, 367–378.

McAlpine, D. D., & Wilson, A. R. (2004). Screening for depression in primary care: What do we still need to know? *Depression and Anxiety, 19*, 137-145.

McCue, P., Buchanan, T., & Martin, C. R. (2006). Screening for psychological distress using internet administration of the Hospital Anxiety and Depression Scale (HADS) in individuals with chronic fatigue syndrome. *British Journal of Clinical Psychology, 45*, 483-498.

McEvoy, P. M., Grove, R., & Slade, T. (2011). Epidemiology of anxiety disorders in the Australian general population: Findings of the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian & New Zealand Journal of Psychiatry, 45*, 957-967.

McGee, S. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine, 17*, 647-650.

McLaughlin, K. A., Gadermann, A. M., Hwang, I., Sampson, N. A., Al-Hamzawi, A., Andrade, L. H., … Kessler, R. C. (2012). Parent psychopathology and offspring mental disorders: Results from the WHO World Mental Health Surveys. *The British Journal of Psychiatry, 200*, 290-299.

Mehlenbacher, B. (2002). Assessing the usability of on-line instructional materials. *New Directions for Teaching and Learning, 91*, 91-98.

Menard, S. (1995) *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Meyer, G. J. (2003). Guidelines for reporting information in studies of diagnostic test accuracy: The STARD initiative. *Journal of Personality Assessment, 81*, 191-193.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., … Read, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128-165.

Michie, S., Smith, J. A., Senior, V., & Marteau, T. M. (2003). Understanding why negative genetic test results sometimes fail to reassure. *American Journal of Medical Genetics Part A, 119*, 340-347.

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage.

Miller, P. R. (2001). Inpatient diagnostic assessments: 2. Interrater reliability and outcomes of structured vs. unstructured interviews. *Psychiatry Research, 105*, 265-271.

Miller, P. R., Dasher, R., Collins, R., Griffiths, P., & Brown, F. (2001). Inpatient diagnostic assessments: Accuracy of structured vs. unstructured interviews. *Psychiatry Research, 105*, 255-264.

Mitchell, A. J., Vaze, A., & Rao, S. (2009). Clinical diagnosis of depression in primary care: A meta-analysis. *The Lancet, 374*, 609-619.

Moos, R. H., Nichol, A. C., Moos, B. S. (2002). Risk factors for symptom exacerbation among treated patients with substance use disorders. *Addiction, 97*, 75-85.

Morgan, D. L. (1993). Qualitative content analysis: A guide to paths not taken. *Qualitative Health Research, 1*, 112-121.

Morgan, J. F., Reid, F., & Lacey, J. H. (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *British Medical Journal, 319*, 1467-1467.

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep, 34*, 601-608.

Morrell, R. W., Mayhorn, C. B., & Bennett, J. H. (2000). A survey of World Wide Web use in middle aged and older adults. *Factors and Ergonomics Society, 42*, 175-182.

Moscicki, E. K. (2001). Epidemiology of completed and attempted suicide: Toward a framework for prevention. *Clinical Neuroscience Research, 1*, 310-323.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old Issues. *American Psychologist, 59*, 150-162.

Naragon-Gainey, K., & Watson, D. (2011). The anxiety disorders and suicidal ideation: Accounting for co-morbidity via underlying personality traits. *Psychological Medicine, 41*, 1437-1447.

Newman, M. (2000). Recommendations for a cost-offset model of psychotherapy allocation using generalised anxiety disorder as an example. *Journal of Consulting and Clinical Psychology, 68*, 549–555.

Neuman, Y., Cohen, Y., Assaf, D., & Kedma, G. (2012). Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine, 56*, 19-25.

Newman, C. F. (1994). Understanding client resistance: Methods for enhancing motivation to change. *Cognitive and Behavioural Practice, 1*, 47-69.

Newman, M. G., Consoli, A., & Taylor, C. B. (1997). Computers in assessment and cognitive behavioral treatment of clinical disorders: Anxiety as a case in point. *Behavior Therapy, 28*, 211-235.

Newman, F. L., Rugh, D., & Ciarlom, J. A. (2004). Guidelines for selecting psychological instruments for treatment planning and outcomes assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Vol. 1. General considerations* (3rd ed., pp. 197-214). Mahwah, NJ: Erlbaum.

Nguyen, D. P. (2013). [Pilot validation of an online clinical diagnostic program for mental disorders]. Unpublished raw data.

Noyes, J.M. & Garland, K.J. (2003). VDT versus paper-based text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics*, *31*, 411-423.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics, 51*, 1352-1375.

Noyes, J. M., Garland, K. J., & Robbins, E. L. (2004). Paper-based versus computer-based assessment: Is workload another test mode effect? *British Journal of Educational Technology*, *35*, 111-113.

Ofrat, S., & Krueger, R. F. (2012). How research on the meta-structure of psychopathology aids in understanding biological correlates of mood and anxiety disorders. *Biology of Mood and Anxiety Disorders, 2*, 13.

Ogles, B. M., France, C. R., Lunnen, K. M., Bell, T., & Goldfarb, M. (1998). Computerized depression screening and awareness. *Community Mental Health Journal, 34*, 27-38.

Ohaeri, J. U. (2002). The burden of caregiving in families with a mental illness: A review of 2002. *Current Opinion in Psychiatry, 16*, 457-465.

Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pretesting survey questions. *Journal of Official Statistics, 7*, 349-365.

Otte, C. (2011). Cognitive behavioral therapy in anxiety disorders: Current state of the evidence. *Dialogues in Clinical Neuroscience, 13*, 413-421.

Otto, M. W., Smits, J. A. J., & Reese, H. E. (2006). Combined psychotherapy and pharmacotherapy for mood and anxiety disorders in adults: Review and analysis. *FOCUS: The Journal of Lifelong Learning in Psychiatry, 4*, 204-214.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing order to the web*. Stanford, CA: Stanford InfoLab.

Parker, G., Fletcher, K., Blanch, B., & Greenfield, L. (2013). The 'real world' utility of a web-based bipolar disorder screening measure. *Acta Psychiatrica Scandinavica, 127*, 373-380.

Paulden, M., Palmer, S., Hewitt, C., & Gilbody, S. (2009). Screening for postnatal depression in primary care: Cost effectiveness analysis. *British Medical Journal, 339*. Retrieved from http://dx.doi.org/10.1136/bmj.b5203

PDM Task Force. (2006). *Psychodynamic Diagnostic Manual*. Silver Spring, MD: Alliance of Psychoanalytic Organizations.

Pedersen, D. M. (1997). Psychological functions of privacy. *Journal of Environmental Psychology, 17*, 147-156.

Peng, C.-Y. J., Lee, K. K., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*, 3-14.

Perlis, R. H. (2005). Misdiagnosis of bipolar disorder. *American Journal of Managed Care, 11*, 271-274.

Peters, L., Clark, D., & Carroll, F. (1998). Are computerized interviews equivalent to human interviewers? CIDI-Auto versus CIDI in anxiety and depressive disorders. *Psychological Medicine, 28*, 893-901.

Petrie, K., & Abell, W. (1994). Responses of parasuicides to a computerized interview. *Computer Human Behaviour, 10*, 415-418.

Phillips, K. A. (2005). *The broken mirror: Understanding and treating body dysmorphic disorder*. New York: Oxford University Press.

Phillips, K. A., Atala, K. D., & Pope, H. G. (1995). *Diagnostic instruments for body dysmorphic disorder.* Paper presented at the New Research Program and Abstracts, American Psychiatric Association 148th annual meeting, Miami.

Phillips, J., Frances, A., Cerullo, M. A., Chardavoyne, J., Decker, H. S., First, M. B., … Zachar, P. (2012). The six most essential questions in psychiatric diagnosis: A pluralogue part 1: conceptual and definitional issues in psychiatric diagnosis. *Philosophy, Ethics, and Humanities in Medicine, 7*, 3.

Pinninti, N. R., Madison, H., Musser, E., & Rissmiller, D. (2003). MINI International Neuropsychiatric Schedule: Clinical utility and patient acceptance. *European Psychiatry, 18*, 361-364.

Pinsonneult, A., & Heppel, N. (1998). Anonymity in group support systems research: New conceptualization, measure, and contingency framework. *Journal of Management Information Systems, 14*(3), 89-108.

Pirkis, J., Harris, M., Hall, W., & Ftanou, M. (2011). *Evaluation of the Better Access to Psychiatrists, Psychologists and General Practitioners through the Medicare Benefits Schedule initiative*. Melbourne, VIC: Centre for Health Policy, Programs, and Economics.

Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of managed care on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment, 70*, 441-447.

Pitt, L., Kilbride, M., Nothard, S., & Morrison, A. P. (2009). Impact of a diagnosis of psychosis: User-led qualitative study. *The Psychiatrist, 33*, 419-423.

Postmes, T., & Spears, R. (2002). Behavior online: Does anonymous computer communication reduce gender inequality? *Personality and Social Psychology Bulletin, 28*, 1073-1083.

Powell, J., & Clarke, A. (2006). Internet information-seeking in mental health: Population survey. *The British Journal of Psychiatry, 189*, 273-277.

Prentice-Dunn, S., & Rogers, R. (1982). Effects of public and private self-awareness deindividuation and aggression. *Journal of Personality and Social Psychology, 43*, 503-513.

Preskorn, S. H., & Baker, B. (2002). The overlap of DSM-IV syndromes: Potential implications for the practice of polypsychopharmacology, psychiatric drug development, and the human genome project. *Journal of Psychiatric Practice, 8*, 170-177.

Pringle, J., Drummond, J., McLafferty, E., & Hendry, C. (2011). Interpretative phenomenological analysis: A discussion and critique. *Nurse Researcher, 18*(3), 20-24.

Prins, M. A., Verhaak, P. F., Bensing, J. M., & van der Meer, K. (2008). Health beliefs and perceived need for mental health care of anxiety and depression: The patients' perspective explored. *Clinical Psychology Review, 28*, 1038-1058.

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

Ramirez-Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendricks, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *The American Journal of Psychiatry*, *157*, 1599–1605.

Rapee, R. M., Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Response to hyperventilation and inhalation of 5.5% carbon dioxide-enriched air across the DSM-III-R anxiety disorders. *Journal of Abnormal Psychology*, *101*, 538-52.

Rash, C., Coffey, S. F., Baschnagel, J. S., Drobes, D., & Saladin, M. E. (2008). Psychometric properties of the IES-R in traumatized substance dependent individuals with and without PTSD. *Addictive Behaviors, 33*, 1039-1047.

Reavley, N. J., & Jorm, A. F. (2011). Recognition of mental disorders and beliefs about treatment and outcome: Findings from an Australian national survey of mental health literacy and stigma. *Australian and New Zealand Journal of Psychiatry*, *45*, 947-956.

Reid, K., Flowers, P., & Larkin, M. (2005). Exploring lived experience. *The Psychologist, 18*, 20-23.

Reid, S. C., Kauer, S. D., Dudgeon, P., Sanci, L. A., Shrier, L. A., & Patton, G. C. (2009). A mobile phone program to track young people's experiences of mood, stress and coping: Development and testing of the mobiletype program. *Social Psychiatry and Psychiatric Epidemiology, 44*, 501-507.

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243-256.

Ramirez-Basco, M., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*, 1599-1605.

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*, 169-184.

Rheingold, H. (1993). *The virtual community: Finding connection in a computerized world*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84*, 754-775.

Riper, H., van Straten, A., Keuken, M., Smit, F., Schippers, G., & Cuijpers, P. (2009). Curbing problem drinking with personalized-feedback interventions: a meta-analysis. *American Journal of Preventative Medicine, 36*, 247-255.

Riper, H., Spek, V., Boon, B., Conijn, B., Kramer, J., Martin-Abello, K., & Smith, F. (2011). Effectiveness of e-self-help interventions for curbing adult problem drinking: A meta-analysis. *Journal of Medical Internet Research, 13*(2), e42.

Ritterband, L. M., Thorndike, F. P., Cox, D. J., Kovatchev, B. P., & Gonder-Frederick, L. A. (2009). A behavior change model for internet interventions. *Annals of Behavioral Medicine, 38*, 18-27.

Riva, G. (1998). Virtual reality as assessment tool in psychology. *Studies in Health Technology and Informatics, 44*, 71-79.

Rizvi, S. L., Peterson, C. B., Crow, S. J., & Agras, W. S. (2000). Test-retest reliability of the eating disorder examination. *International Journal of Eating Disorders, 28*, 311-316.

Rizzo, A., Parsons, T. D., Kenny, P., & Buckwalter, J. G. (2012). Using virtual reality for clinical assessment and intervention. In L. L'Abate & D. A. Kaiser (Eds.), *Handbook of Technology in Psychology, Psychiatry and Neurology*. Hauppause, New York: Nova Science Publishers.

Roberts, M., & Ilardi, S. S. (2008). *Handbook of research methods in clinical psychology*. Retrieved from http://SWIN.eblib.com.au/patron/FullRecord.aspx?p=214148

Robins, L. N. (1994). How to choose among the riches: Selecting a diagnostic instrument. *International Review of Psychiatry, 6*, 265-271.

Robins, L. N., & Cottler, L. B. (2004). Making a structured psychiatric diagnostic interview faithful to the nomenclature. *American Journal of Epidemiology, 160*, 808-813.Robins, L. N., & Guze, S. B. (1975). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry, 126*, 983-987.

Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry, 38*, 381-389.

Robinson, R., & West, R. (2002). A comparison of computers and questionnaire methods of history-taking in a genito-urinary clinic. *Psychology and Health, 6*, 77-84.

Roca-Bennasar, M., Garcia-Mas, A., Llaneras, N., & Blat, J. (1991). Kraepelin: An expert system for the diagnosis of obsessive-compulsive disorders. *European Psychiatry 6*, 171-175.

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1997). Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *American Journal of Psychiatry, 154*, 1593-1598.

Roid, G. H., & Gorsuch, R. L. (1984). Development and clinical use of test-interpretive programs on microcomputers. In M. D. Schwartz (Ed.), *Using computers in clinical practice: Psychotherapy and mental health applications* (pp. 141-149). New York: Haworth Press.

Rome, H. P., Swenson, W. M., Mataya, P., McCarthy, C. E., Pearson, J. S., Keating, F. R., & Hathaway, S. R. (1962). *Symposium on automation technics in personality assesment.* Paper presented at the Staff Meetings of the Mayo Clinic.

Rosenfeld, P., Booth-Kewley, S., Edwards, J. E., & Thomas, M. D. (1996). Responses on computer surveys: Impression management, social desirability, and the big brother syndrome. *Computers in Human Behavior, 12*, 263-274.

Rosenman, S. J., Korten, A. E., & Levings, C. T. (1997). Computerised diagnosis in acute psychiatry: Validity of CIDI-Auto against routine clinical diagnosis. *Journal of Psychiatric Research, 31*, 581-592.

Roshanaei-Moghaddam, B., Pauly, M. C., Atkins, D. C., Baldwin, S. A., Stein, M. B., & Roy-Byrne, P. (2011). Relative effects of CBT and pharmacotherapy in depression versus anxiety: Is medication somewhat better for depression, and CBT somewhat better for anxiety? *Depression and Anxiety, 28*, 560-567.

Rotondi, A. J., Sinkule, J., Haas, G. L., Spring, M. B., Litschge, C. M., Newhill, C. E., … Anderson, C. M. (2007). Designing websites for persons with cognitive deficits: Design and usability of a psychoeducational intervention for persons with severe mental illness. *Psychological Services, 4*, 202-224.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 394-404.

Ruland, C. M., Bakken, S., & Røislien, S. (2007). Reliability and validity issues related to interactive tailored patient assessments. *Journal of Medical Internet Research, 9*(3), 1-9.

Russell, E.W. (1995). The accuracy of automated and clinical detection of brain damage and lateralization in neuropsychology. *Neuropsychology Review*, *5*, 1-68.

Rush, J., First, M. B., & Blacker, D. (2008). *Handbook of psychiatric measures*. Washington, DC: American Psychiatric Press Inc.

Ruwaard, J., Lange, A., Broeksteeg, J., Renteria-Agirre, A., Schrieken, B., Dolan, C. V., & Emmelkamp, P. (2012). Online cognitive-behavioural treatment of bulimic symptoms: A randomized controlled trial. *Clinical Psychology and Psychotherapy, 39*, 443-456.

Ryan, J. (2010). *A history of the internet and the digital future*. London: Reakton Books Ltd.

Sale, R. (2006). International guidelines on computer-based and internet-delivered testing: A practitioner's perspective. *International Journal of Testing, 6*, 181-188.

SANE. (2007). Research Bulletin 4: Stigma and mental illness.

Santor, D. A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in depression. *Measurement, 4*, 135-155.

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption. *Addiction, 88*, 791-804.

Scarvalone, P. A., Cloitre, M., Spielman, L. A., Jacobsberg, L., Fishman, B., & Perry, S. W. (1996). Distress reduction during the Structured Clinical Interview for DSM-III-R. *Psychiatry Research, 59*, 245-249.

Scheff, T. J. (1974). The labelling theory of mental illness. *American Sociological Review*, *39*, 444-452.

Schmidt, W. C. (1997). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments & Computers, 29*, 274-279.

Schulenberg, S. E., & Yutrzenka, B. A. (2001). Equivalence of computerized and conventional versions of the Beck Depression Inventory-II (BDI-II). *Current Psychology, 20*, 216.

Schulze, B. (2007). Stigma and mental health professionals: A review of the evidence on an intricate relationship. *International Review of Psychiatry, 19*, 137-155.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation, 22*, 127-160.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.

Shear, M. K., Bjelland, I., Beesdo, K., Gloster, A. T., & Wittchen, H.-U. (2007). Supplementary dimensional assessment in anxiety disorders. *International Journal of Methods in Psychiatric Research, 16*, S52-S64.

Shear, M. K., Rucci, P., Williams, J., Frank, E., Grochocinski, V., Vander Bilt, J., … Wang, T. (2001). Reliability and validity of the Panic Disorder Severity Scale: Replication and extension. *Journal of Psychiatric Research, 35*, 293-296.

Sheatsley, P. B. (1983). Questionnaire construction and item writing. In P. H. Rossi, J. D. Wright & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 195-230). New York: Academic Press.

Sheehan, D. V., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., … Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry, 12*, 232-241.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Supplement 20), 22-33.

Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton: Chapman & Hall.

Sim, K., Gwee, K. P., Bateman, A. (2005). Case formulation in psychotherapy: Revitalizing its usefulness as a clinical tool. *Academic Psychiatry*, *29*, 289-292.Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy 85*, 257–268.

Slade, T., & Andrews, J. (2002). Exclusion criteria in the diagnostic classifications of DSM-IV and ICD-10: Revisiting the co-occurrence of psychiatric syndromes. *Psychological Medicine, 32*, 1203-1211.

Slade, T., Johnston, A., Teesson, M., Whiteford, H., Burgess, P., Pirkis, J., & Saw, S. (2009). *The mental health of Australians 2. Report on the 2007 National Survey of Mental Health and Wellbeing.* Canberra: Department of Health and Ageing.

Slade, T., & Watson, D. (2006). The structure of common DSM–IV and ICD–10 mental disorders in the Australian general population. *Psychological Medicine, 36*, 1593-1600.

Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal of Quality in Health Care, 18*, 377-382.

Smit, N., Smit, F., Cuijpers, P., & De Graaf, R. (2007). Using decision theory to derive optimal cut-off scores of screening instruments: An illustration explicating costs and benefits of mental health screening. *International Journal of Methods in Psychiatric Research, 16*, 219-229.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: Theory, method, and research*. London, UK: Sage Publications.

Smith, J. A., & Osborn, M. (2008). Interpretative phenomenological analysis. In J. A. Smith (Ed.), *Qualitative psychology - A practical guide to research methods* (2nd ed.). Wiltshire, UK: Sage.

Somers, J. M., Goldner, E. M., Waraich, P., & Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: A systematic review of the literature. *Canadian Journal of Psychiatry, 51*, 100-113.

Spek, V., Cuijpers, P., Riper, H., Keyzer, J., & Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: A meta-analysis. *Psychological Medicine, 37*, 319-328.

Spek, V., Nyklícek, I., Cuijpers, P., & Pop, V. (2008). Internet administration of the Edinburgh Depression Scale. *Journal of Affective Disorders, 106*, 301-305.

Spiers, N., Bebbington, P., McManus, S., Brugha, T. S., Jenkins, R., & Meltzer, H. (2011). Age and birth cohort differences in the prevalence of common mental disorder in England: the National Psychiatric Morbidity Surveys. 1993-2007. *British Journal of Psychiatry 198*, 479.

Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry, 24*, 399-411.

Spitzer, R. L., Endicott, J., & Robins, E. (1975). Clinical criteria for psychiatric diagnoses and DSM-III. *American Journal of Psychiatry*, *132,* 1187-1192.

Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry, 125*, 341-347.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*, 1092-1097.

Spitzer, R. L., Williams, J. B. R., Kroenke, K., Linzer, M., deGruy, F. V., Hahn, S. R., … Johnson, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care:

The PRIME-MD 1000 study. *The Journal of the American Medical Association, 272*, 1749-1756.

Spitzer, R. L., & Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? *American Journal of Psychiatry, 156*, 1856-1864.

Sproull, L, & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

Stieger, S., & Reips, U. -D. (2008) Dynamic Interviewing Program (DIP): Automatic online interviews via the instant messenger ICQ. *Cyberpsychology & Behaviour*, *11*, 201-207.

Stein, M. B., Walker, J. R., & Forde, D. R. (1994). Setting diagnostic thresholds for social phobia: Considerations from a community survey of social anxiety. *American Journal of Psychiatry, 151*, 408-412.

Stengler-Wenzke, K., Beck, M., Holzinger, A., & Angermeyer, M. C. (2004). Stigma experiences of patients with obsessive compulsive disorders. *Fortschritte Neurologie-Psychiatrie, 72*, 7-13.

Steketee, G., Frost, R., & Bogart, K. (1996). The Yale-Brown Obsessive Compulsive Scale: Interview versus self-report. *Behaviour Research and Therapy, 34*, 675–684.

Stewart, S. E., Ceranoqlu, T. A., O'Hanley, T., & Geller, D. A. (2005). Performance of clinician versus self-report measures to identify obsessive-compulsive disorder in children and adolescents. *Journal of Child and Adolescence Psychopharmacology, 15*, 956-963.

Stewart, R. E., & Chambless, D. L. (2009). Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consultative Clinical Psychology, 77*, 595-606.

Stinchfield, R. (2002). Reliability, validity, and classification accuracy of the South Oaks Gambling Screen (SOGS). *Addictive Behaviors, 27*, 1-19.

Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment, 81*, 209-219.

Substance Abuse and Mental Health Services Administration. (2010). *Results from the 2009 National Survey on Drug Use and Health: Mental health findings*. Rockville, MD: Office of Applied Studies, NSDUH.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behaviour*, *7*, 321-326.

Suler, J., Barak, A., Chechele, P., Fenichel, M., Hsiung, R., Maguire, J., Meunier, V., … Walker-Schmucker, W. (2001). Assessing a person's suitability for online therapy: The ISMHO clinical case study group. *CyberPsychology & Behavior, 4*, 675-679.

Summerfeldt, L. J., Klooserman, P. H., & Antony, M. M. (2010). Structured and semi-structured diagnostic interviews. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 95-140). New York: The Guilford Press.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Tarrier, N., Liversidge, T., & Gregg, L. (2006). The acceptability and preference for the psychological treatment of PTSD. *Behaviour Research and Therapy, 44*, 1643-1656.

Tate, D. F., Finkelstein, E. A., Khavjou, O., & Gustafon, A. (2009). Cost effectiveness of internet interventions: Review and recommendations. *Annals of Behavioural Medicine, 38*, 40-45.

Tate, D. F., & Zabinski, M. F. (2004). Computer and Internet applications for psychological treatment: Update for clinicians. *Journal of Clinical Psychology, 60*, 209-220.

Taylor, H. (1999). Explosive growth of “Cyberchondriacs” continues. *The Harris Poll #47, 5 August 1999*. Retrieved from http://www.harrisinteractive.com

Taylor, S., Abramowitz, J. S., & McKay, D. (2010). Obsessive-compulsive disorder. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 267-300). New York: The Guilford Press.

Taylor, S. T. & Bogdan, R. (1998). Introduction to qualitative research methods: A guidebook and resource (3rd ed.). New York: John Wiley & Sons.

Taylor, C. B., Fried, L., & Kenardy, J. (1990). The use of a real-time computer diary for data acquisition and processing. *Behaviour Research and Therapy, 21*, 93-97.

Taylor, C. B., & Luce, K. H. (2003). Computer- and internet-based psychotherapy interventions. *Current Directions in Psychological Science, 12*, 18-22.

Thase, M. E., & Sachs, G. S. (2000). Bipolar depression: Pharmacotherapy and related therapeutic strategies. *Biological Psychiatry, 48*, 558-572.

Tiemens, B. G., Ormel, J., & Simon, G. E. (1996). Occurrence, recognition, and outcome of psychological disorders in primary care. *American Journal of Psychiatry, 153*, 636-644.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*, 275-304.

Tullis, T., & Albert, B. (2008). *Measuring the user experience collecting, analyzing, and presenting usability metrics*. Boston MA: Elsevier.

Turner, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report, 15*, 754-760.

Umefjord, G., Petersson, G., & Hamberg, K. (2003). Reasons for consulting a doctor on the internet: Web survey of users of an Ask the Doctor service. *Journal of Medical Internet Research, 5*(4), e26.

Ustun, D., & Kennedy, C. (2009). What is "functional impairment"? Disentangling disability from clinical significance. *World Psychiatry, 8*, 82-85.

Vallejo, M. A., Jordan, C. M., Diaz, M., Comeche, M., & Ortega, J. (2007). Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the general health questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Tevised (SLR-90-R). *Journal of Medical Internet Research, 9*(1), 1-10.

Van Ameringen, M., Mancini, C., Simpson, W., & Patterson, B. (2010). Potential use of Internet-based screening for anxiety disorders: A pilot study. *Depression & Anxiety, 27*, 1006-1010.

Van Beljouw, M. I., Verhaak, P., Prins, M. M., Cuijpers, P., Penninx, B., & Bensing, J. (2010). Reasons and determinants for not receiving treatment for common mental disorders. *Psychiatric Services, 61*, 250-257.

Van der Feltz-Cornelis, C. M., Knispel, A., & Elfeddali, I. (2008). Treatment of mental disorder in the primary care setting in the Netherlands in the light of the new reimbursement system: A challenge? *International Journal of Integrated Care, 8*, e56.

Van der Krieke, L., Emerencia, A. C., Aiello, M., & Sytema, S. (2012). Usability evaluation of a web-based support system for people with a schizophrenia diagnosis. *Journal of Medical Internet Research, 14*(1), e24.

Van der Linden, W. J., & Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics, 29*, 273-291.

Van Rijswijk, E., van Hout, H., van de Lisdonk, E., Zitman, F., & van Weel, C. (2009). Barriers in recognising, diagnosing and managing depressive and anxiety disorders as experienced by family physicians: A focus group study. *BMC Family Practice, 10*, 52.

Van Straten, A., Seekles, W., van't Veer-Tazelaar, N. J., Beekman, A. T. F., & Cuijpers, P. (2010). Stepped care for depression in primary care: What should be offered and how? *Medical Journal of Australia, 192*(11), 36.

Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F. & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Quality of Life Research, 16*(Suppliment 1), 143-155.

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 43-66). Hoboken, New Jersey: Wiley.

Weber, B., Fritze, J., Schneider, B., Kühner, T., & Maurer, K. (2002). Bias in computerized neuropsychological assessment of depressive disorders caused by computer attitude. *Acta Psychiatrica Scandinavica, 105*, 126-130.

Webster, J., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behavior Research Methods, Instruments & Computers, 28*, 567-576.

Weiller, E., Bisserbe, J. C., Boyer, P., Lepine, J. P., & Lecrubier, Y. (1996). Social phobia in general health care: An unrecognised undertreated disabling disorder. *British Journal of Psychiatry, 168*, 169-174.

Weisband, S., & Kiesler, S. (1996). *Self-disclosure on computer forms: Meta analysis and implications.* Paper presented at the CHI '96 Conference on Human Factors in Computing Systems, Vancouver, BC.

Weiss, D. S., & Marmar, C. R. (1996). The Impact of Event Scale - Revised. In J. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD*. New York: Guilford.

Weldon, L. J., Koved, L., & Shneiderman, B. (1985). *The structure of information in online and paper technical manuals.* Paper presented at the 29th annual meeting of the Human Factors Society, Santa Monica, CA.

Wichers, M., Simons, C. J. P., Kramer, I. M. A., Hartmann, J. A., Lothmann, C., Myin-Germeys, I., …van Os, J. (2011). Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica, 124*, 262-272.

Widiger, T. A., & Trull, T. (1991). Diagnosis and clinical assessment. *Annural Review of Psychology, 42*, 109-133.

Williams, R., Bambling, M., King, R., & Abbott, Q. (2009). In-session processes in online counselling with young people: An exploratory approach. *Counselling and Psychotherapy Research*, *9*(2), 93-100.

Winters, K. C., Specker, S., Stinchfield, R. D. (1997). *Diagnostic Interview for Gambling Severity (DIGS)*. Minneapolis: University of Minnesota.

Wilson, M. (1993). DSM-III and the transformation of American psychiatry: a history. *American Journal of Psychiatry, 150*, 399-410.

Winzelberg, A. (1997). The analysis of an electronic support group for individuals with eating disorders. *Computers in Human Behaviour, 13*, 393-407.

Wittchen, H. U., Lachner, G., & Wunderlich, U. (1998). Test-retest reliability of the computerised DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Social Psychiatry and Psychiatric Epidemiology, 33*, 568-578.

Wittchen, H. U., & Pittrow, D. (2002). Prevalence, recognition and management of depression in primary care in Germany: The Depression 2000 study. *Human Psychopharmacology, 17*(Supplement 1), 1-11.

Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology, 53*, 519-543.

World Health Organization. (1993). *Composite International Diagnostic Interview - Version 1.1*. Geneva: WHO.

World Health Organisation International Consortium in Psychiatric Epidemiology. (2000). Cross-national comparisons of the prevalences and correlates of mental disorder. *Bulletin of the World Health Organisation, 78*, 413-426.

World Health Organisation. (2008). *The global burden of disease: 2004 update*. Geneva: World Health Organisation.

Yonkers, K. A., Bruce, S. E., Dyck, I. R., & Keller, M. B. (2003). Chronicity, relapse, and illness-course of panic disorder, social phobia, and generalized anxiety disorder: Findings in men and women from 8 years of follow-up. *Depression and Anxiety, 17*, 173-179.

Yoshino, A., Shigemura, J., Kobayashi, Y., Nomura, S., Shishikura, K., … Ashida, H. (2001). Telepsychiatry: Assessment of televideo psychiatric interview reliability with present- and next-generation internet infrastructures. *Acta Psychiatrica Scandinavica*, *104*, 223-226.

Yu, S.-C., & Yu, M.-N. (2007). Comparison of Internet-based and paper-based questionnaires in Taiwan using multisample invariance approach. *CyberPsychology & Behavior, 10*, 501-507.

Zbozinek, T. D., Rose, R. D., Wolitzky-Taylor, K. B., Sherbourne, C., Sullivan, G., Stein, M. B., … Craske, M. G. (2012). Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. *Depression & Anxiety, 29*, 1065-1071.

Zich, J. M., Attkisson, C. C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: The CES-D and the BDI. *International Journal of Psychiatry in Medicine, 20*, 259-277.

Ziefle, M., 1998. Effects of display resolution on visual performance. *Human Factors*, *40*, 554-568.

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order vs. Deindividuation, impulse and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska symposium on motivation* (Vol. 17, pp. 237–307). Lincoln, NE: University of Nebraska Press.

Zimmerman, M., & Mattia, J. I. (1999). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry, 40*, 182-191.

Appendices

Appendix A: Ethics clearance

Below is a copy of an email from the Swinburne University Human Research Ethics Committee (dated 10/2/2009) indicating ethics clearance for the research conducted in this thesis.

SUHREC Project 2008/143 - The evaluation of the psychometric properties and user experience of an internet-based clinical assessment schedule
Assoc Prof Britt Klein, FLSS; Mr David Phong Nguyen
Approved Duration: 10/02/2008 to 31/12/2010

Dear Assoc Prof Britt Klein and Mr David Nguyen

I refer to the ethical review of the above project protocol undertaken by Swinburne's Human Research Ethics Committee (SUHREC). Your responses to the review, as emailed on 6 January 2009 with attachments, were put to the Committee's delegate and feedback sent to you. Further clarification/revision emailed today appears in line with the delegate's feedback.

I am pleased to advise that, as submitted to date, the project has approval to proceed in line with standard on-going ethics clearance conditions here outlined.

- *All human research activity undertaken under Swinburne auspices must conform to Swinburne and external regulatory standards, including the National Statement on Ethical Conduct in Human Research and with respect to secure data use, retention and disposal.*
- *The named Swinburne Chief Investigator/Supervisor remains responsible for any personnel appointed to or associated with the project being made aware of ethics clearance conditions, including research and consent procedures or instruments approved. Any change in chief investigator/supervisor requires timely notification and SUHREC endorsement.*
- *The above project has been approved as submitted for ethical review by or on behalf of SUHREC. Amendments to approved procedures or instruments ordinarily require prior ethical appraisal/ clearance. SUHREC must be notified immediately or as soon as possible thereafter of a) any serious or unexpected adverse effects on participants and any redress measures; b) proposed changes in protocols; and (c) unforeseen events which might affect continued ethical acceptability of the project.*
- *At a minimum, an annual report on the progress of the project is required as well as at the conclusion (or abandonment) of the project.*
- *A duly authorised external or internal audit of the project may be undertaken at any time.*

Please contact me if you have any queries about on-going ethics clearance. The SUHREC project number should be quoted in communication.

Best wishes for the project.
Yours sincerely
Keith Wilkins
Secretary, SUHREC, Research Ethics Officer, Swinburne Research (H68)
Swinburne University of Technology

Appendix B: STARD checklist

Section and Topic	Item	
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend heading 'sensitivity and specificity').
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.
METHODS		
Participants	3	Describe the study population: The inclusion and exclusion criteria, setting and locations where the data were collected.
	4	Describe participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the (evaluated) index tests or the (golden) reference standard?
	5	Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.
	6	Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	Describe the reference standard and its rationale.
	8	Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.
	9	Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.
	10	Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.
	11	Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).
	13	Describe methods for calculating test reproducibility, if done.
RESULTS		
Participants	14	Report when study was done, including beginning and ending dates of recruitment.
	15	Report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, co morbidity, current treatments, recruitment centers).
	16	Report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).
Test results	17	Report time interval from the index tests to the reference standard, and any treatment administered between.
	18	Report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.
	20	Report any adverse events from performing the index tests or the reference standard.
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).
	22	Report how indeterminate results, missing responses and outliers of the index tests were handled.
	23	Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.
	24	Report estimates of test reproducibility, if done.
DISCUSSION	25	Discuss the clinical applicability of the study findings.

Appendix C: Information sheet and consent form

Evaluating the e-Psychological Assessment (e-PASS) Program

Principal Investigator: Assoc. Professor Britt Klein, Co-Director: National eTherapy Centre
Student Investigator: David Nguyen, Research Assistant: National eTherapy Centre
Associate Investigator: Assoc. Professor David Austin, Co-Director: National eTherapy Centre

e-PASS

The electronic Psychological Assessment (e-PASS) program is an online clinical assessment tool for identifying common psychological disorders. Developed by the Swinburne eTherapy Unit, e-PASS is the first publicly available online clinical assessment program developed within Australia and it is free to use on the Anxiety Online website. By replying to a series of questions presented by e-PASS, a respondent can find out whether they may have a psychological disorder, and if so, its severity. So far, e-PASS has not been formally tested to show that it produces accurate and consistent feedback or can perform similar to a formal clinical assessment. As a result, e-PASS users are currently recommended to see a mental health professional should they require a full clinical assessment.

Research Project Aims

This project aims to evaluate the accuracy and reliability of e-PASS and how users perceive the experience of undertaking e-PASS. Previous research suggests that simple online assessment programs can be effective in identifying certain psychological disorders. However, less known is whether more comprehensive programs such as e-PASS can accurately and reliably assess for a range of psychological disorders and with varying severity. Furthermore, there has been little research to date exploring the user experience of an online assessment program such as e-PASS in comparison to a traditional assessment interview with a mental health professional. If e-PASS is shown to be accurate, reliable and well accepted by users, it may prove to be a highly valuable tool for health professionals and the public.

Participation Requirements

To participate in this project, you are required to be 18 years or older and reside in Australia.

1. If you choose to participate, you will be asked to immediately complete the e-PASS, which will take roughly 5-30 minutes.
2. You will then be notified by email to complete an online questionnaire about psychological symptoms you may have initially reported. This will take approximately 15 minutes to complete.
3. Approximately one week after completing the e-PASS, you will be asked to undergo a clinical interview over the telephone. There is the option of undertaking the interview in-person at the

Swinburne Psychology Clinic (in Hawthorn, Victoria) if convenient for you. The interview will last approximately 45 minutes and can be arranged to occur during a time of your availability.

4. Shortly after completing the e-PASS and interview, you will be invited to complete an online survey about your experience of the two activities and this will last roughly 10 minutes.
5. Finally, a small number of participants (which may include you) will be asked to complete the e-PASS for a second time.

To participate, you are required to supply your contact phone number and email address at the time of consenting to participate in this project. We also ask that you supply the name and phone number of your GP in case you experience an emergency during participation, and we require notifying your GP.

Participation is entirely voluntary and you are free to withdraw at any time, for any reason. If you require, you can also have your data removed during or after the course of your participation. The completion of the following consent form will be taken as your consent to participate.

Should I participate if I'm feeling very unwell?

Participation will involve answering personal questions about certain psychological symptoms you may have. While this can be beneficial and therapeutic, it may also be a confronting and distressing experience if you are feeling particularly distressed or vulnerable. Therefore, if you are currently feeling suicidal or going through a crisis, it is recommended that you delay participating in this project until you feel better, and immediately contact a trained counsellor at Lifeline on 13 11 14 (if you are in Australia) or a crisis support service in your area for professional assistance. For a list of phone and internet resources across Australia, please click here.

If you experience distress as a result of participating, you are strongly recommended to contact Lifeline (11 13 14 – available 24/7 across Australia) and to talk to your GP or a mental health professional such as a registered psychologist. The clinical interview which all participants undertake should not be used for counselling.

Privacy and confidentiality

Your privacy and confidentiality will be treated as matters of utmost importance. All online interaction for this project will be done through a secure website with considerable security measures similar to those used by online banking websites. The interview will be conducted with confidentiality and participants are advised to undertake the telephone interview in a private setting. All participant data will be stored in an encrypted database only accessible by the investigators. Should you participate, your contact details (email and telephone number) will be stored separately from all other data and will be deleted once you have finished participating. An identification code will be used to label the various sets of data collected from your participation. Participant data will be retained for 5 years after the completion of this project and then destroyed.

Who is conducting this research?

This project is conducted by the Swinburne eTherapy Unit. The student investigator is undertaking the project as part of their research requirements in completing a Doctorate in Clinical Psychology.

Will the results of this research be published?

The results of the study are expected to be published in the student investigator's research thesis and may be published in an academic journal article. Only group data will be reported, hence no individual participant will be identifiable. By completing the following consent form, you will be giving consent for your data to be used in the study and for publication.

This project has been approved by Swinburne University's Human Research Ethics Committee in line with the National Statement on Ethical Conduct in Human Research. Any queries regarding this research can be directed to Assoc. Professor Britt Klein at bklein@swin.edu.au or 03 9214 8851.

If you have any concerns or complaints about the conduct of this project, you can contact:
Research Ethics Officer, Swinburne Research (H68),
Swinburne University of Technology, P O Box 218, HAWTHORN VIC 3122.
Tel (03) 9214 5218 or +61 3 9214 5218 or resethics@swin.edu.au

Your participation is very important to the study and will be greatly appreciated. Thank you for your time and assistance.

If you intend to participate in this research project, it is highly recommended that you keep a copy of this information sheet. Please print out a copy of this page. For a downloadable PDF copy, please click here.

Appendix C: Information sheet and consent form (continued)

Evaluating the Validity and Reliability of the e-Psychological Assessment (e-PASS) Program

Swinburne University of Technology
Project Title: Evaluating the Validity and Reliability of the online Psychological Assessment (e-PASS) Program
Investigators: Assoc. Professor Britt Klein, David Nguyen, and Assoc. Professor David Austin

1. I consent to participate in the project named above. I have been provided a copy of the project information statement and this consent form and any questions I have asked have been answered to my satisfaction.
2. By providing consent to participate in this research:
    - I confirm that I am 18 years or older and live in Australia
    - I agree to perform the e-Psychological Assessment (e-PASS) program
    - I agree to complete an online psychological questionnaire
    - I agree to be interviewed over the telephone (or in person at the Swinburne Psychology Clinic if agreed) approximately one week after completing the e-PASS. I agree to allow my telephone interview to be tape-recorded for quality purposes, if randomly selected and on the condition that I am notified beforehand.
    - I agree to respond to the online survey of user experience after I complete the e-PASS and interview.
    - I agree to allow my e-PASS, interview, user experience survey, and follow-up psychological questionnaire results to be collected as data for this research project
    - I agree to make myself available to undertake the e-PASS a second time if asked to by email
3. By consenting to participate in this research, I acknowledge that:
    - The potential adverse effects of participation have been explained to me to my satisfaction;
    - My participation is voluntary and that I am free to withdraw from the project at any time without explanation;
    - The project is for the purpose of research and not for profit;
    - My contact details will be used by the investigators to communicate with me in relation to my participation in this project
    - By supplying my GP's contact details below, I permit the investigators to notify my GP in the event that I experience a crisis or emergency situation during my participation in this project
    - Any personal or health information about me which is gathered in the course of and as the result of my participating in this project will be (i) collected and retained for the purpose of this project and (ii) accessed and analysed by the project investigators for the purpose of conducting this project;
    - This project may be published in the student investigator's research thesis and/or in a journal article. In both cases, my anonymity will be preserved and I will not be identified in publications or otherwise without my expressed written consent.

By entering my name, my email address, contact phone number, preferred contact time, GP contact details, and clicking the 'Continue' button below, I agree to participate in this project.

Appendix D: Research advertisment

The following was presented to Anxiety Online website users who clicked on a link to access the e-PASS:

The National eTherapy Centre invites you to participate in research evaluating the Online Psychological Assessment (e-PASS) program. To be eligible, you must be 18 years or older and live in Australia. Participation is voluntary and will involve completing the e-PASS, a telephone interview, an online questionnaire about psychological symptoms, and an online survey about your experience of the e-PASS. Please click here if you would like to find out more about this research project.

Appendix E: Clinical interview form

EPASS EVALUATION - CLINCAL INTERVIEW FORM				
Interviewer: Participant initials: Recorded Phone Call: Y N Final 3 digits of telephone number: Date of interview: Length of interview (mins):				
SEVERITY SCALE				
Use this scale when discussing symptom severity with participants: 0 (absent) - 2 (Mild) - 4 (Moderate) - 6 (Severe) - 8 (very Severe)				
MINI:	**Participant Rating:** **(including ADIS and other schedules)** (out of 8)	**Clinician Rating** **(including ADIS and other schedules)** (out of 8)	**Additional Notes** (including differential diagnosis/further information required etc)	**Confidence in diagnosis** (low/med/high)
Suicidality (according to MINI) Low Mod High* **Comments** (* If high, please assess risk in more detail, refer to appropriate resources and report to David and Britt for supervision)				
Confirmed ePASS diagnosis? Yes No	**Comments**			
Do they agree with ePASS results? Scale: 1(Completely disagree) to 5 (Completely agree)	**Comments:**			

Appendix F: Clinical questionnaires

Panic Disorder Severity Scale (PDSS)

ePASS - Followup Questionnaires

PDSS

The following section is about **panic attacks**.A panic attack is a feeling of fear or apprehension that begins suddenly and builds rapidly in intensity, usually reaching a peak inless than 10 minutes. This feeling is associated with uncomfortablephysical sensations like racing or pounding heart, shortness of breath,choking, dizziness, sweating, trembling. Often there are distressing,catastrophic thoughts such as fear of losing control, having a heartattack or dying. A panic attack has *at least four* such symptoms.

For the following statements, select the response that best represents your response to that statement. Please make sure that you answer every item.

2. Frequency of panic attacks

- (•) None
- ( ) Mild (panic sensations or limited symptoms or less than one full panic attack per week)
- ( ) Moderate (one or two full panic attacks per week)
- ( ) Severe (daily panic attacks or more than two a week)
- ( ) Extreme (attacks occur more than once a day)

3. Distress during panic attacks:

- (•) None
- ( ) Mild (infrequent and not too intense)
- ( ) Moderate (regular and intense but still manageable)
- ( ) Severe (very frequent and very intense)
- ( ) Extreme (distress with all attacks)

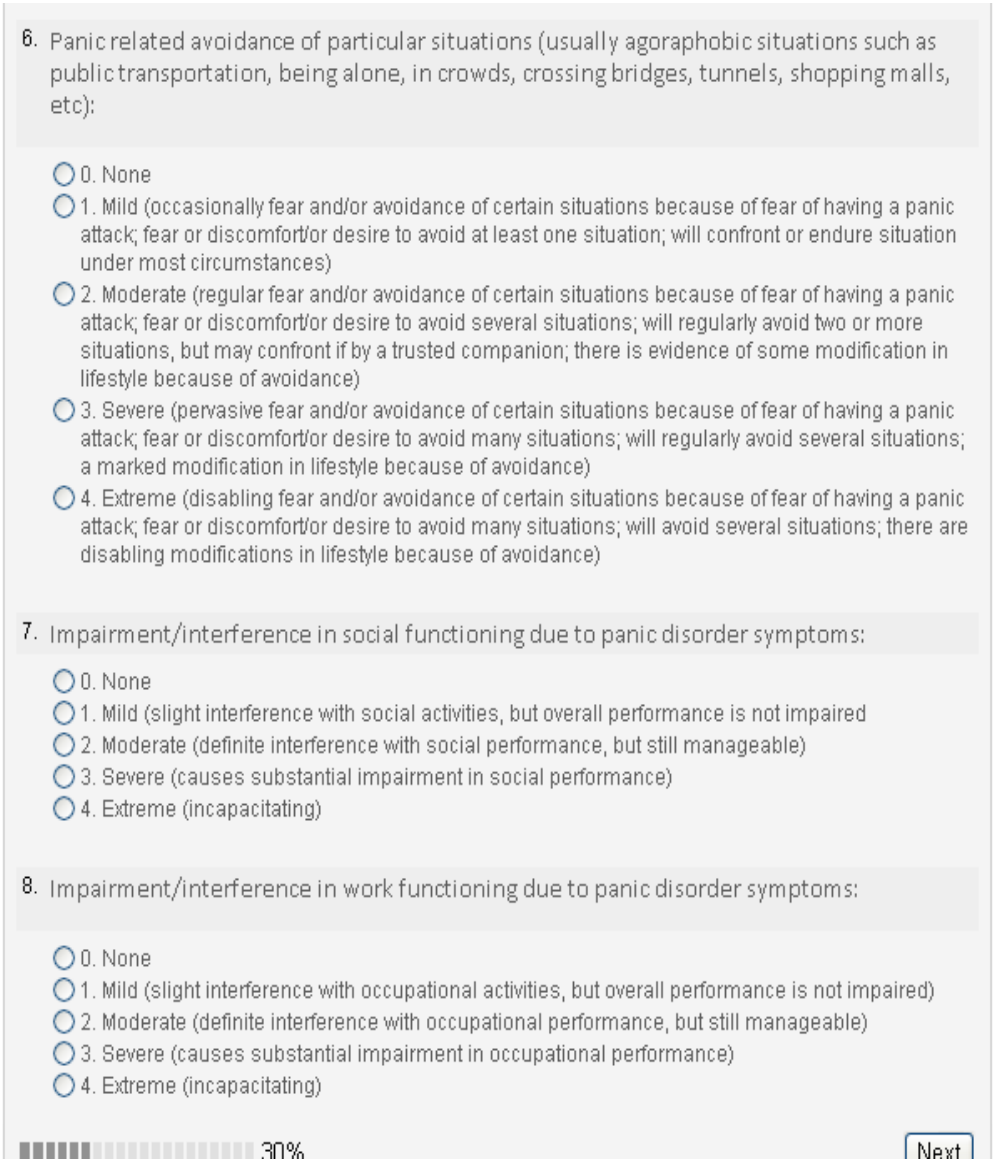
4. Anticipatory anxiety (worry about future panic attacks):

- ( ) None
- ( ) Mild (occasionally worry about when the next panic will occur)
- ( ) Moderate (frequent worry about next attack)
- ( ) Severe (preoccupied with very disturbing worry about next attack)
- ( ) Extreme (near constant and disabling worry)

5. Panic related phobic avoidance ofsensations (eg. avoid exercise, avoid expressing strong emotions, etc):

- ( ) 0. None
- ( ) 1. Mild (occasionally fear and/or avoidance of physical sensations because of fear of having a panic attack; fear or discomfort with one or more physical sensations; will endure sensations under most circumstances)
- ( ) 2. Moderate (regular fear and/or avoidance of physical sensations because of fear of having a panic attack; fear or discomfort/or desire to avoid several physical sensations; has reduced some activities to limit sensations)
- ( ) 3. Severe (pervasive fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort with experiencing sensations and/or avoidance leads to marked constriction of lifestyle)
- ( ) 4. Extreme (disabling fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort/or desire to avoid many activities; there are disabling modifications in lifestyle because of avoidance)

Panic Disorder Severity Scale (PDSS; continued)



6. Panic related avoidance of particular situations (usually agoraphobic situations such as public transportation, being alone, in crowds, crossing bridges, tunnels, shopping malls, etc):

- ○ 0. None
- ○ 1. Mild (occasionally fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort/or desire to avoid at least one situation; will confront or endure situation under most circumstances)
- ○ 2. Moderate (regular fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort/or desire to avoid several situations; will regularly avoid two or more situations, but may confront if by a trusted companion; there is evidence of some modification in lifestyle because of avoidance)
- ○ 3. Severe (pervasive fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort/or desire to avoid many situations; will regularly avoid several situations; a marked modification in lifestyle because of avoidance)
- ○ 4. Extreme (disabling fear and/or avoidance of certain situations because of fear of having a panic attack; fear or discomfort/or desire to avoid many situations; will avoid several situations; there are disabling modifications in lifestyle because of avoidance)

7. Impairment/interference in social functioning due to panic disorder symptoms:

- ○ 0. None
- ○ 1. Mild (slight interference with social activities, but overall performance is not impaired
- ○ 2. Moderate (definite interference with social performance, but still manageable)
- ○ 3. Severe (causes substantial impairment in social performance)
- ○ 4. Extreme (incapacitating)

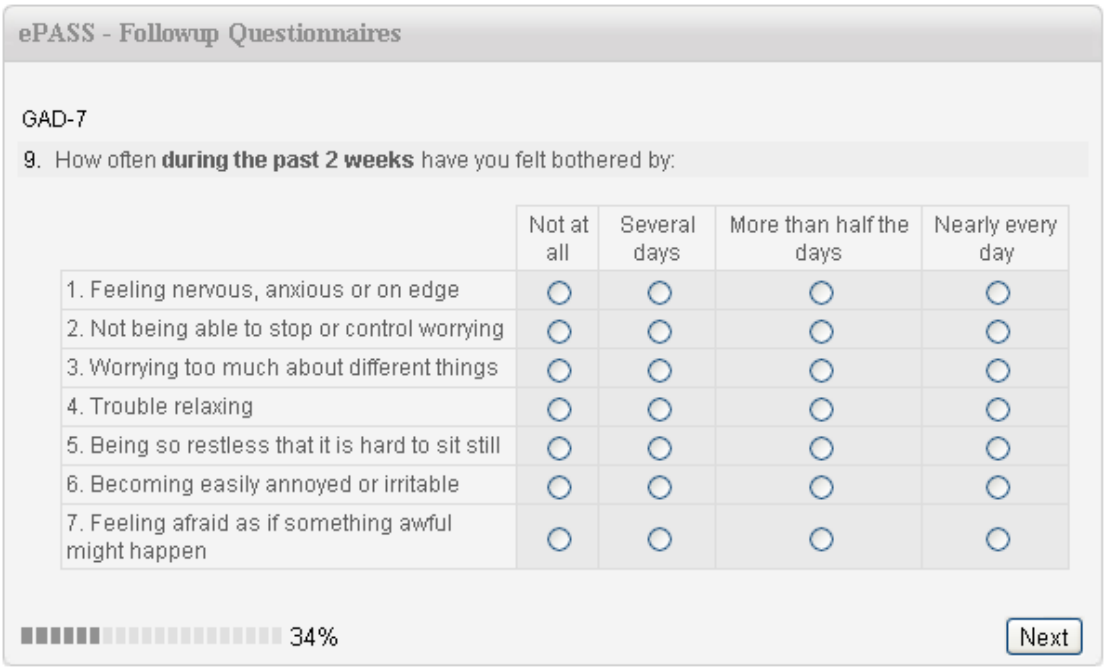
8. Impairment/interference in work functioning due to panic disorder symptoms:

- ○ 0. None
- ○ 1. Mild (slight interference with occupational activities, but overall performance is not impaired)
- ○ 2. Moderate (definite interference with occupational performance, but still manageable)
- ○ 3. Severe (causes substantial impairment in occupational performance)
- ○ 4. Extreme (incapacitating)

30%

Next

Generalised Anxiety Disorder-7 (GAD-7)



ePASS - Followup Questionnaires

GAD-7

9. How often **during the past 2 weeks** have you felt bothered by:

	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious or on edge	○	○	○	○
2. Not being able to stop or control worrying	○	○	○	○
3. Worrying too much about different things	○	○	○	○
4. Trouble relaxing	○	○	○	○
5. Being so restless that it is hard to sit still	○	○	○	○
6. Becoming easily annoyed or irritable	○	○	○	○
7. Feeling afraid as if something awful might happen	○	○	○	○

34%

Next

Centre for Epidemiologic Studies Depression Scale (CES-D)

CES-D

10. Below is a list of the ways you might have felt or behaved. Please tell me how often you have felt this way during the past week

	Rarely or none of the time (less than 1 day)	Some or a little of the time (1-2 days)	Occasionally or a moderate amount of time (3-4 days)	Most or all of the time (5-7 days)
1. I was bothered by things that usually don't bother me.	○	○	○	○
2. I did not feel like eating; my appetite was poor.	○	○	○	○
3. I felt that I could not shake off the blues even with help from my family or friends.	○	○	○	○
4. I felt I was just as good as other people.	○	○	○	○
5. I had trouble keeping my mind on what I was doing.	○	○	○	○
6. I felt depressed.	○	○	○	○
7. I felt that everything I did was an effort.	○	○	○	○
8. I felt hopeful about the future.	○	○	○	○
9. I thought my life had been a failure.	○	○	○	○
10. I felt fearful.	○	○	○	○
11. My sleep was restless.	○	○	○	○
12. I was happy.	○	○	○	○
13. I talked less than usual.	○	○	○	○
14. I felt lonely.	○	○	○	○
15. People were unfriendly.	○	○	○	○
16. I enjoyed life.	○	○	○	○
17. I had crying spells.	○	○	○	○
18. I felt sad.	○	○	○	○
19. I felt that people dislike me.	○	○	○	○
20. I could not get going.	○	○	○	○

38%

Next

Social Phobia Inventory (SPIN)

SPIN

11. Please check how much the following problems have bothered you during the past week. Mark only one box for each problem, and be sure to answer all items.

	Not at all	A little bit	Somewhat	Very much	Extremely
1. I am afraid of people in authority.	○	○	○	○	○
2. I am bothered by blushing in front of people.	○	○	○	○	○
3. Parties and social events scare me.	○	○	○	○	○
4. I avoid talking to people I don't know.	○	○	○	○	○
5. Being criticized scares me a lot.	○	○	○	○	○
6. Fear of embarrassment causes me to avoid doing things or speaking to people.	○	○	○	○	○
7. Sweating in front of people causes me distress.	○	○	○	○	○
8. I avoid going to parties.	○	○	○	○	○
9. I avoid activities in which I am the center of attention.	○	○	○	○	○
10. Talking to strangers scares me.	○	○	○	○	○
11. I avoid having to give speeches.	○	○	○	○	○
12. I would do anything to avoid being criticized.	○	○	○	○	○
13. Heart palpitations bother me when I am around people.	○	○	○	○	○
14. I am afraid of doing things when people might be watching.	○	○	○	○	○
15. Being embarrassed or looking stupid are my worst fears.	○	○	○	○	○
16. I avoid speaking to anyone in authority.	○	○	○	○	○
17. Trembling or shaking in front of others is distressing to me.	○	○	○	○	○

42%

Next

Yale-Brown Obsessive Compulsive Scale –Self-Report (Y-BOCS-SR)

YALE-BROWN OBSESSIVE COMPULSIVE SCALE (Y-BOCS)*

Questions 1 to 5 are about your obsessive thoughts.

Obsessions are unwanted ideas, images or impulses that intrude on thinking against your wishes and efforts to resist them. They usually involve themes of harm, risk and danger. Common obsessions are excessive fears of contamination; recurring doubts about danger; extreme concern with order, symmetry, or exactness; fear of losing important things.

Please answer each question by writing the appropriate number in the box next to it.

1. TIME OCCUPIED BY OBSESSIVE THOUGHTS

Q. How much of your time is occupied by obsessive thoughts?

0 = None.
1 = Less than 1 hr/day or occasional occurrence.
2 = 1 to 3 hrs/day or frequent.
3 = Greater than 3 and up to 8 hrs/day or very frequent occurrence.
4 = Greater than 8 hrs/day or nearly constant occurrence.

2. INTERFERENCE DUE TO OBSESSIVE THOUGHTS

Q. How much do your obsessive thoughts interfere with your work, school, social, or other important role functioning? Is there anything that you don't do because of them?

0 = None.
1 = Slight interference with social or other activities, but overall performance not impaired.
2 = Definite interference with social or occupational performance, but still manageable.
3 = Causes substantial impairment in social or occupational performance.
4 = Incapacitating.

3. DISTRESS ASSOCIATED WITH OBSESSIVE THOUGHTS

Q. How much distress do your obsessive thoughts cause you?

0 = None.
1 = Not too disturbing.
2 = Disturbing, but still manageable.
3 = Very disturbing.
4 = Near constant and disabling distress.

4. RESISTANCE AGAINST OBSESSIONS

Q. How much of an effort do you make to resist the obsessive thoughts? How often do you try to disregard or turn your attention away from these thoughts as they enter your mind?

0 = Try to resist all the time.
1 = Try to resist most of the time.
2 = Make some effort to resist.
3 = Yield to all obsessions without attempting to control them, but with some reluctance.
4 = Completely and willingly yield to all obsessions.

5. DEGREE OF CONTROL OVER OBSESSIVE THOUGHTS

Q. How much control do you have over your obsessive thoughts? How successful are you in stopping or diverting your obsessive thinking? Can you dismiss them?

0 = Complete control.
1 = Usually able to stop or divert obsessions with some effort and concentration.
2 = Sometimes able to stop or divert obsessions.
3 = Rarely successful in stopping or dismissing obsessions, can only divert attention with difficulty.
4 = Obsessions are completely involuntary, rarely able to even momentarily alter obsessive thinking.

*This adaptation of the Y-BOCS is abridged from the original version with permission from Wayne Goodman. For additional information on the Y-BOCS, please contact Dr. Wayne Goodman at the University of Florida, College of Medicine, Gainesville, Florida 32610. The original version was published by: Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale I: Development, use, and reliability. *Arch Gen Psychiatry*. 1989;46:1006-1011.

Y-BOCS-SR (continued)

The next several questions are about your compulsive behaviors.

Compulsions are urges that people have to do something to lessen feelings of anxiety or other discomfort. Often they do repetitive, purposeful, intentional behaviors called rituals. The behavior itself may seem appropriate but it becomes a ritual when done to excess. Washing, checking, repeating, straightening, hoarding and many other behaviors can be rituals. Some rituals are mental. For example thinking or saying things over and over under your breath.

6. TIME SPENT PERFORMING COMPULSIVE BEHAVIORS

Q. How much time do you spend performing compulsive behaviors? How much longer than most people does it take to complete routine activities because of your rituals? How frequently do you do rituals?

☐
0 = None.
1 = Less than 1 hr/day, or occasional performance of compulsive behaviors.
2 = From 1 to 3 hrs/day, or frequent performance of compulsive behaviors.
3 = More than 3 and up to 8 hrs/day, or very frequent performance of compulsive behaviors.
4 = More than 8 hrs/day, or near constant performance of compulsive behaviors (too numerous to count).

7. INTERFERENCE DUE TO COMPULSIVE BEHAVIORS

Q. How much do your compulsive behaviors interfere with your work, school, social, or other important role functioning? Is there anything that you don't do because of the compulsions?

☐
0 = None.
1 = Slight interference with social or other activities, but overall performance not impaired.
2 = Definite interference with social or occupational performance, but still manageable.
3 = Causes substantial impairment in social or occupational performance.
4 = Incapacitating.

8. DISTRESS ASSOCIATED WITH COMPULSIVE BEHAVIOR

Q. How would you feel if prevented from performing your compulsion(s)? How anxious would you become?

☐
0 = None.
1 = Only slightly anxious if compulsions prevented.
2 = Anxiety would mount but remain manageable if compulsions prevented.
3 = Prominent and very disturbing increase in anxiety if compulsions interrupted.
4 = Incapacitating anxiety from any intervention aimed at modifying activity.

9. RESISTANCE AGAINST COMPULSIONS

Q. How much of an effort do you make to resist the compulsions?

☐
0 = Always try to resist.
1 = Try to resist most of the time.
2 = Make some effort to resist.
3 = Yield to almost all compulsions without attempting to control them, but with some reluctance.
4 = Completely and willingly yield to all compulsions.

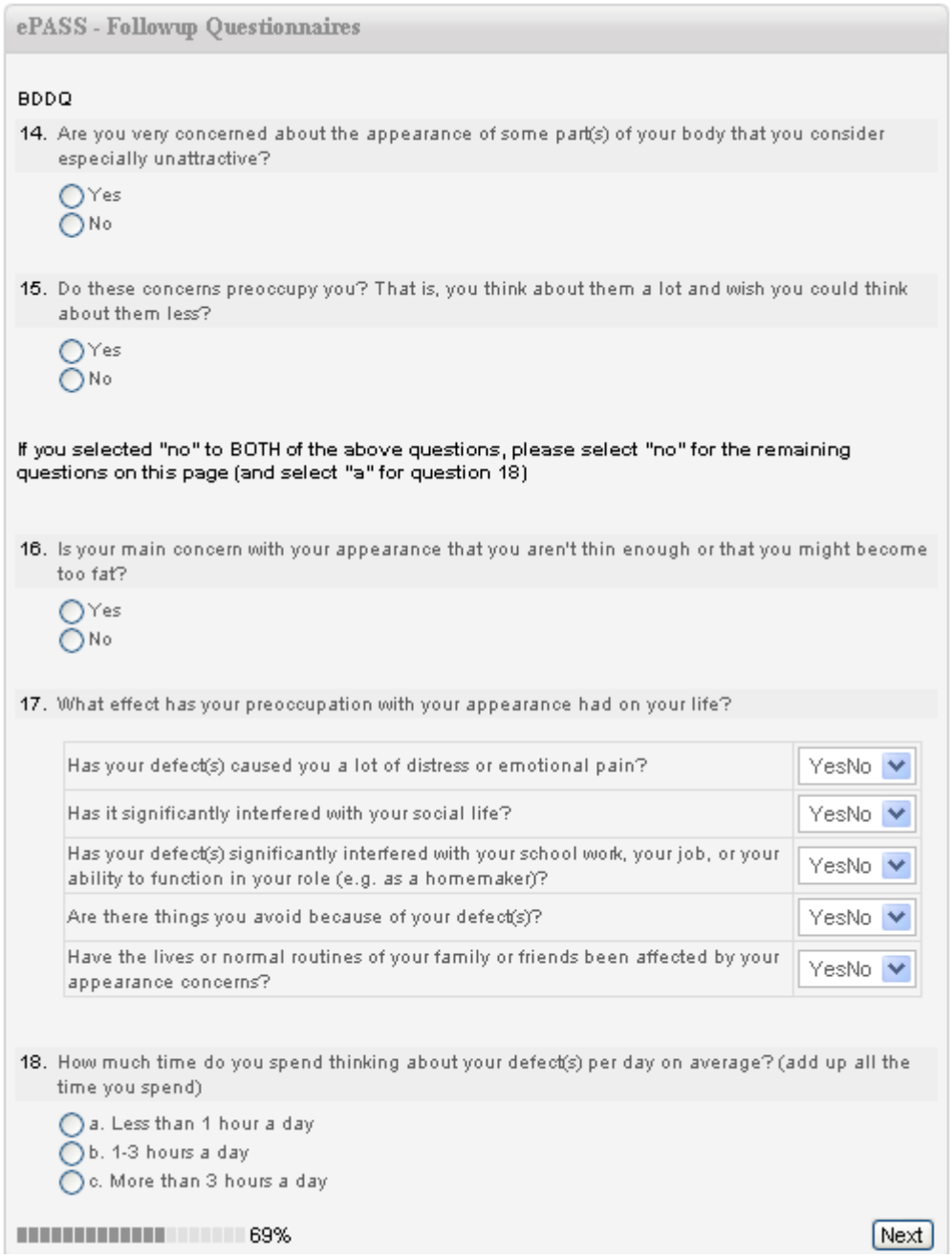
10. DEGREE OF CONTROL OVER COMPULSIVE BEHAVIOR

Q. How strong is the drive to perform the compulsive behavior? How much control do you have over the compulsions?

☐
0 = Complete control.
1 = Pressure to perform the behavior but usually able to exercise voluntary control over it.
2 = Strong pressure to perform behavior, can control it only with difficulty.
3 = Very strong drive to perform behavior, must be carried to completion, can only delay with difficulty.
4 = Drive to perform behavior experienced as completely involuntary and overpowering, rarely able to even momentarily delay activity.

☐ **Total Score**

Body Dysmorphic Disorder Questionnaire (BDDQ)



ePASS - Followup Questionnaires

BDDQ

14. Are you very concerned about the appearance of some part(s) of your body that you consider especially unattractive?

- ◯ Yes
- ◯ No

15. Do these concerns preoccupy you? That is, you think about them a lot and wish you could think about them less?

- ◯ Yes
- ◯ No

If you selected "no" to BOTH of the above questions, please select "no" for the remaining questions on this page (and select "a" for question 18)

16. Is your main concern with your appearance that you aren't thin enough or that you might become too fat?

- ◯ Yes
- ◯ No

17. What effect has your preoccupation with your appearance had on your life?

Has your defect(s) caused you a lot of distress or emotional pain?	YesNo
Has it significantly interfered with your social life?	YesNo
Has your defect(s) significantly interfered with your school work, your job, or your ability to function in your role (e.g. as a homemaker)?	YesNo
Are there things you avoid because of your defect(s)?	YesNo
Have the lives or normal routines of your family or friends been affected by your appearance concerns?	YesNo

18. How much time do you spend thinking about your defect(s) per day on average? (add up all the time you spend)

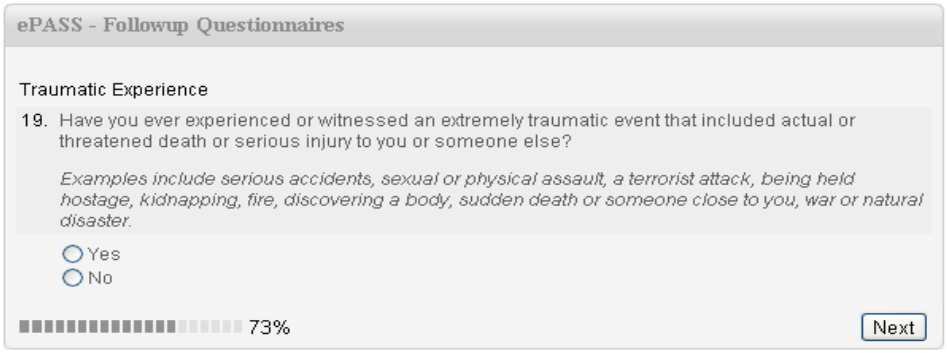
- ◯ a. Less than 1 hour a day
- ◯ b. 1-3 hours a day
- ◯ c. More than 3 hours a day

69%

Next

Impact of Event Scale – Revised (IES-R)

Initial screening question to determine whether to present the IES-R



IES-R

IES-R

20. Below is a list of difficulties people sometimes have after very stressful (or traumatic) life events (e.g. an incident where either you or someone close to you faced considerable danger).

If you have previously experienced a stressful life event, please read each item, and then indicate how distressing each difficulty has been for you during the past seven days. Otherwise, select '0' for all of the items on this page.

	0 - Not at all	1 - A little bit	2 - Moderately	3 - Quite a bit	4 - Extremely
Any reminder brought back feelings about it.	○	○	○	○	○
I had trouble staying asleep.	○	○	○	○	○
Other things kept making me think about it.	○	○	○	○	○
I felt irritable and angry.	○	○	○	○	○
I avoided letting myself get upset when I thought about it or was reminded of it.	○	○	○	○	○
I thought about it when I didn't mean to.	○	○	○	○	○
I felt as if it hadn't happened or wasn't real.	○	○	○	○	○
I stayed away from reminders about it.	○	○	○	○	○
Pictures about it popped into my mind.	○	○	○	○	○
I was jumpy and easily startled.	○	○	○	○	○
I tried not to think about it.	○	●	○	○	○
I was aware that I still had a lot of feelings about it, but I didn't deal with them.	○	○	○	○	○
My feelings about it were kind of numb.	○	○	○	○	○
I found myself acting or feeling like I was back at that time.	○	○	○	○	○
I had trouble falling asleep.	○	○	○	○	○
I had waves of strong feelings about it.	○	○	○	○	○
I tried to remove it from my memory.	○	○	○	○	○
I had trouble concentrating.	○	○	○	○	○
Reminders of it caused me to have physical reactions, such as sweating, trouble breathing, nausea, or a pounding heart.	○	○	○	○	○
I had dreams about it.	○	○	○	○	○
I felt watchful and on guard.	○	○	○	○	○
I tried not to talk about it.	○	○	○	○	○

Fear Questionnaire (FQ)

ePASS - Followup Questionnaires

FQ

21. Choose a number from the scale below to show how much you would avoid each of the situations listed below because of fear or other unpleasant feelings.

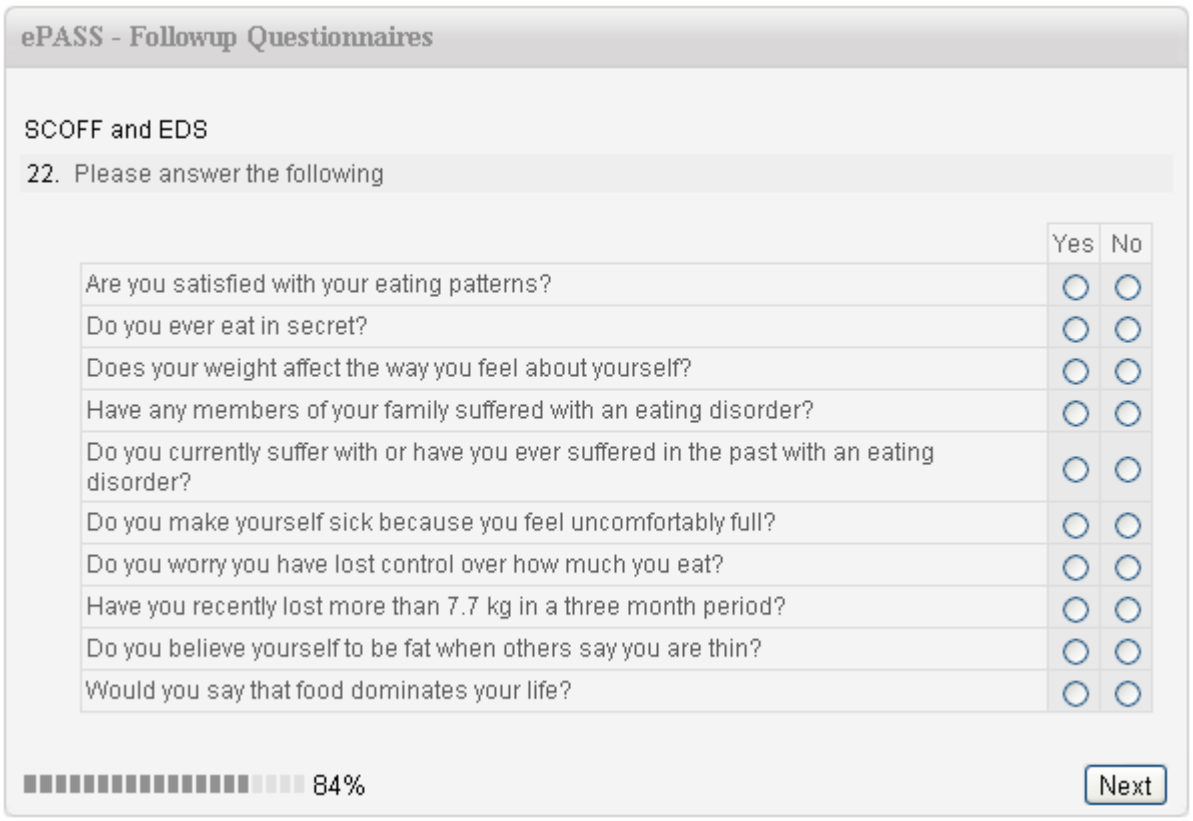
0	1	2	3	4	5	6	7	8
Would not avoid it		Slightly avoid it		Definitely avoid it		Markedly avoid it		Always avoid it

	0	1	2	3	4	5	6	7	8
Injections or minor surgery	○	○	○	○	○	○	○	○	○
Eating or drinking with other people	○	○	○	○	○	○	○	○	○
Hospitals	○	○	○	○	○	○	○	○	○
Travelling alone or by bus	○	○	○	○	○	○	○	○	○
Walking alone in busy streets	○	○	○	○	○	○	○	○	○
Being watched or stared at	○	○	○	○	○	○	○	○	○
Going into crowded shops	○	○	○	○	○	○	○	○	○
Talking to people in authority	○	○	○	○	○	○	○	○	○
Sight of blood	○	○	○	○	○	○	○	○	○
Being criticised	○	○	○	○	○	○	○	○	○
Going alone far from home	○	○	○	○	○	○	○	○	○
Thought of injury or illness	○	○	○	○	○	○	○	○	○
Speaking or acting to an audience	○	○	○	○	○	○	○	○	○
Large open spaces	○	○	○	○	○	○	○	○	○
Going to the dentist	○	○	○	○	○	○	○	○	○
Animals (e.g. dogs, cats, snakes)	○	○	○	○	○	○	○	○	○
Natural events (e.g. lightning, storms, earthquakes)	○	○	○	○	○	○	○	○	○
Body fluids (e.g. faeces, vomit, semen)	○	○	○	○	○	○	○	○	○
Materials (e.g. cleaning products, medicine, poison)	○	○	○	○	○	○	○	○	○
Specific situations (e.g. driving, riding elevators, crossing bridges)	○	○	○	○	○	○	○	○	○
Other situations (e.g. items, places, people) - please briefly describe		○	○	○	○	○	○	○	○	○
Other situations (e.g. items, places, people) - please briefly describe		○	○	○	○	○	○	○	○	○

80%

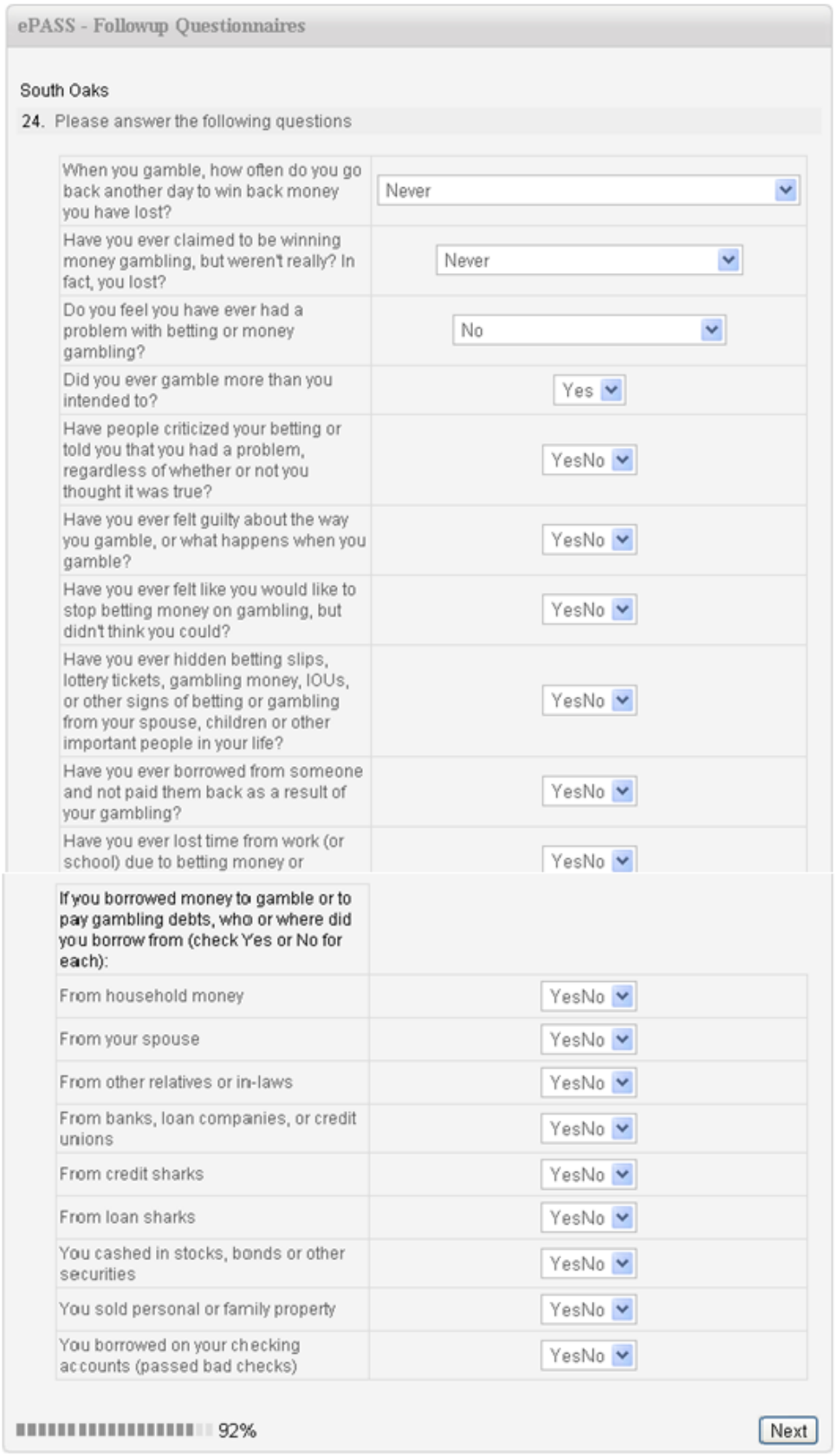
Next

SCOFF and Eating Disorder Screen for Primary Care (ESP)



Gambling screening question

South Oaks Gambling Scale



ePASS - Followup Questionnaires

South Oaks

24. Please answer the following questions

When you gamble, how often do you go back another day to win back money you have lost?	Never
Have you ever claimed to be winning money gambling, but weren't really? In fact, you lost?	Never
Do you feel you have ever had a problem with betting or money gambling?	No
Did you ever gamble more than you intended to?	Yes
Have people criticized your betting or told you that you had a problem, regardless of whether or not you thought it was true?	YesNo
Have you ever felt guilty about the way you gamble, or what happens when you gamble?	YesNo
Have you ever felt like you would like to stop betting money on gambling, but didn't think you could?	YesNo
Have you ever hidden betting slips, lottery tickets, gambling money, IOUs, or other signs of betting or gambling from your spouse, children or other important people in your life?	YesNo
Have you ever borrowed from someone and not paid them back as a result of your gambling?	YesNo
Have you ever lost time from work (or school) due to betting money or	YesNo

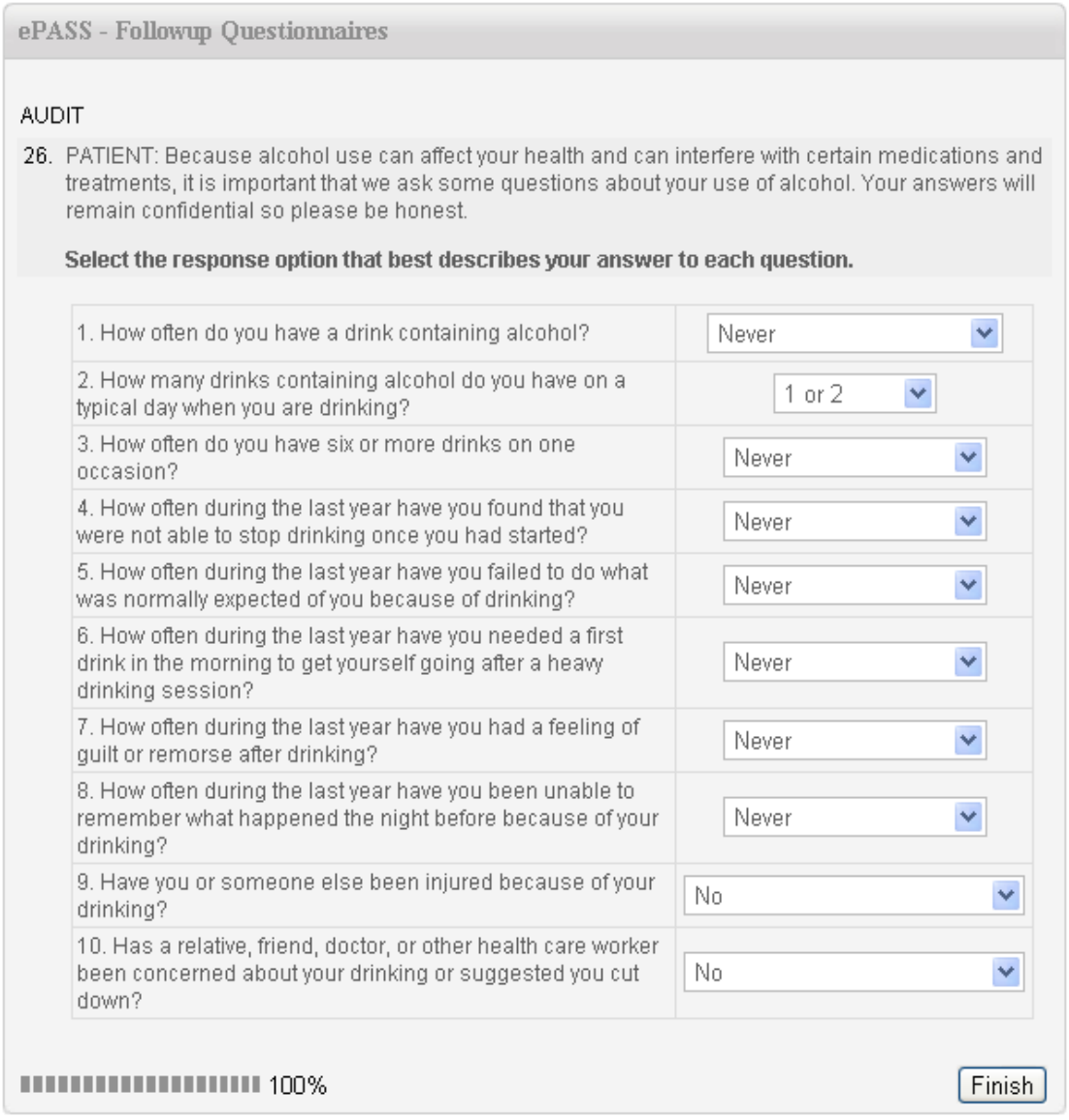
If you borrowed money to gamble or to pay gambling debts, who or where did you borrow from (check Yes or No for each):	
From household money	YesNo
From your spouse	YesNo
From other relatives or in-laws	YesNo
From banks, loan companies, or credit unions	YesNo
From credit sharks	YesNo
From loan sharks	YesNo
You cashed in stocks, bonds or other securities	YesNo
You sold personal or family property	YesNo
You borrowed on your checking accounts (passed bad checks)	YesNo

92%

Next

Alcohol screening question

Alcohol Use Disorders Identification Test (AUDIT)



Appendix G: List of abbreviations

ADIS	Anxiety Disorders Interview Schedule
AO	Anxiety Online
APA	American Psychiatric Association
AUC	Area Under the Curve
BDD	Body dysmorphic disorder
CBT	Cognitive behavioural therapy
CI	Clinical Interview
DSM	Diagnostic and Statistical Manual
CIDI	Composite International Diagnostic Interview
e-PASS	Online Psychological Assessment program
H-L	Hosmer-Lemeshow statistic
GAD	Generalised anxiety disorder
K-6	Kessler-6
LL	Log likelihood statstic
LRM	Logistic regression model
MDD	Major depressive episode
MINI	Mini International Neuropsychiatric Interview
NLR	Negative likelihood ratio
NPV	Negative predictive value
OCD	Obsessive-compulsive disorder
PCA	Principle component analysis
PLR	Positive likelihood ratio
p-p	Paper and pencil
PPV	Positive predictive value
PTSD	Post-traumatic stress disorder
ROC	Receiver Operating Characteristic
SCID	Structured Clinical Interview for DSM Disorders
UE	User experience
WHO	World Health Organisation