RELATIONSHIPS AMONG COGNITION, SYMPTOM VALIDITY, AND SELF-REPORTED DISABILITY

_____

A Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

_____

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

_____

By

W. Alexander Alverson, M.A.

August, 2017

RELATIONSHIPS AMONG COGNITION, SYMPTOM VALIDITY, AND SELF-REPORTED DISABILITY

_____
W. Alexander Alverson, M.A.

APPROVED:

_____
Paul Massman, Ph.D.
Committee Co-Chair
University of Houston Department of Psychology

_____
Robert Collins, Ph.D., ABBP-CN
Committee Co-Chair
Michael E. DeBakey Veterans Affairs Medical Center / Baylor College of Medicine

_____
Steven Woods, PsyD
University of Houston Department of Psychology

_____
Matthew Gallagher, Ph.D.
University of Houston Department of Psychology

_____
Antonio D. Tillis, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of Hispanic Studies

Acknowledgements

To my advisor, mentor, and great supporter, Dr. Paul Massman, whose enthusiasm for neuropsychology is only matched by his passion for the success of his students, thank you immensely for your endless support. You helped introduce me to neuropsychology and instilled in me a great respect for the field. I look forward to telling my future trainees of the great legacy I have been a part of.

To one of my first clinical supervisors, Dr. Rob Collins, thank you for introducing me to the VA and epilepsy, challenging me to think critically, and for being an incredible supporter during confusing and trying times of internship and fellowship applications, interviews, and decision-making. You met me at my level of training, pushed me to become a better psychologist, and helped me build the confidence and competence I needed to be successful.

To Dr. Steven Woods, thank you for your brilliant ideas and excellent feedback. I hope to bring your natural talent for creative inquiry to all of my future research endeavors. To Dr. Matt Gallagher, for helping me find the information I needed to make this a successful project, and being supportive and flexible through this process, thank you.

To my friends for life, the Neuronauts, Elyssa Gerst, Marika Faytell, and Lee Wiegand, thank you for being there, for the support, the laughs, and the late night study sessions. When we were all new to Houston and new to grad school, we were there for each other, and will be always. I am proud to have you as friends.

To my family, I know you are as happy for me to finish school as I am, and I promise I’m not going back after this. Thank you for always pushing me, always

believing in me, always keeping me on track, and setting strong examples for me to follow. You are all role models to me, and I love you all very much.

To my wife, Damjana, for bearing the invisible burden of having a spouse in a doctoral program for five years, with unwavering support and poise, and through the hardship of physical distance during my internship year, you have my unconditional and everlasting admiration, gratitude, and love. As we approach the end of this long chapter of our lives, I could not be happier to have you by my side as we begin the next. You have always been my biggest supporter, and I could not have done it without you; I wouldn't have wanted to.

RELATIONSHIPS AMONG COGNITION, SYMPTOM VALIDITY, AND SELF-REPORTED DISABILITY

_____

An Abstract of a Dissertation

Presented to

The Faculty of the Department

of Psychology

University of Houston

_____

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

_____

By

W. Alexander Alverson, M.A.

August, 2017

Abstract

Objective: Disability and functioning are central aspects of neuropsychological and psychological evaluations. Disability is often assessed by self-report measures, such as the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0), which are only marginally related to objective cognitive functioning, and more strongly related to depression and psychiatric symptom severity. In many settings, the accuracy of a patient's responding (symptom validity) may be called into question, particularly when there is an identified or possible external incentive. It was hypothesized that symptom validity would moderate the relationship of cognition and disability, and that failure of a symptom validity test (SVT), the MMPI-2-RF Infrequent Somatic Complaints scale, would be associated with greater self-reported disability.

Methods: This study examined the interrelationships among depression, cognition, symptom validity, and self-reported disability in a sample of Veterans undergoing evaluation for seizure disorders at the Michael E. DeBakey VA Medical Center Epilepsy Monitoring Unit. Structural equation modeling approaches were used to examine the extent to which depression, cognition, and symptom validity predict self-reported disability. Follow-up analyses including means comparisons and chi square tests were used to detect group differences in symptom endorsement as well as associated demographic factors.

Results: Structural equation modeling analyses indicated that depression was the strongest predictor of self-reported disability, accounting for almost all of the variance explained. There was no evidence of a moderating effect of symptom validity on the relationship between objective cognitive performance and self-reported disability.

Models which excluded depression revealed that symptom validity is moderately predictive of self-reported disability. Subjects who failed SVTs were more likely to report greater disability and symptom severity across self-report instruments. Diagnosis was associated with SVT failure such that patients with psychogenic nonepileptic events were more likely to fail SVT than patients with epilepsy.

Conclusions: The study added to a growing literature on the utility of self-reported disability by providing further evidence that objective cognitive performance is not associated with perceived disability, whereas levels of depression, and to a lesser extent, the tendency to report infrequent symptoms, are more strongly related. Self-report of disability should therefore be interpreted with caution, and in the context of psychiatric factors.

TABLE OF CONTENTS

Acknowledgements iii
Abstract vi
Introduction 1
Measuring Disability Using WHODAS 2.0 5
Measuring disability in the epilepsy monitoring unit (EMU). 8
Cognitive performance in the EMU. 9
Symptom Validity in the EMU. 10
Symptom Validity and Disability. 11
Aims of the current study 12
Aim 1. 12
Aim 2. 13
Hypotheses 13
Aim 1. 13
Aim 2. 14
Methods 14
Participants 14
Measures 15
Disability/Functioning. 15
Depression 17
Symptom Validity. 19
Cognition 19
Performance validity. 22
Analysis 23
Results 27
Data properties and sample characteristics 27
Statistical power. 27
Normality. 28
Missing data. 29
Descriptive statistics. 30

Aim 1 32
Correlations. 32
Measurement model. 34
Structural models. 38
Aim 2 46
Discussion 50
Aim 1 50
Aim 2 52
Contributions of the present study 53
Limitations 54
Future Directions 58
References 60
Appendix 1. WHODAS 2.0, 36-item self-report version 81
Appendix 2. Selected Published WHODAS 2.0 Means Table 83

Introduction

Measuring disability provides clinically meaningful information about a patient's functional limitations and the impact of a disease or disorder on that patient's livelihood beyond what can be learned from diagnosis alone (Üstün, 2010). Assessing the practical implications of health conditions allows clinicians to understand patients in the context of their disease. The World Health Organization (WHO) has established the International Classification of Functioning, Disability, and Health (ICF; World Health Organization, 2001) to provide a biopsychosocial framework for the definition and assessment of disability. Consistent with this framework, WHO has developed the Disability Assessment Schedule 2.0 (WHODAS 2.0) as a standardized and nonspecific measurement tool for the evaluation of functional limitations (Üstün, 2010; Üstün et al., 2010). WHODAS 2.0 is a widely available instrument that has been validated in a broad array of patient populations including mood, psychotic, and chronic pain disorders, stroke, and other chronic health conditions (Federici & Meloni, 2010; Federici, Meloni, & Presti, 2009; Garin et al., 2010; Guilera et al., 2015; McKibbin, Patterson, & Jeste, 2004). Interestingly, no validation studies took place in samples of patients with epilepsy, and very few studies in general have examined WHODAS 2.0 in these patients. This is surprising, considering the occurrence of disability, cognitive problems, and psychiatric comorbidities in these patients (Kessler, Lane, Shahly, & Stang, 2012). The availability and ease of use of the self-report version of WHODAS 2.0 makes it a valuable clinical tool across a range of settings, particularly as ICF defines the construct of disability to be independent of its etiology.

Disability as measured by responses to WHODAS 2.0 items presumes equivalence between reported and actual levels of functioning, however it is generally found that perceived (i.e. self-reported) disability is not strongly correlated with measures of functional capacity (Kaye et al., 2014; McKibbin et al., 2004; Naismith, Longley, Scott, & Hickie, 2007). Still, report of subjective limitations, such as WHODAS 2.0, provides meaningful information about overall functioning, and quality of life (Carlozzi et al., 2015; Hudson, Steele, Taillefer, Baron, & Canadian Scleroderma Research Group, 2008). As such, the relationship of WHODAS 2.0 to other measures of functioning is of great clinical interest and utility. While WHODAS 2.0 shows excellent convergent validity with other self-report measures of functioning (Garin et al., 2010; Küçükdeveci et al., 2013; McKibbin et al., 2004; Pösl, Cieza, & Stucki, 2007; Üstün et al., 2010), and is associated with both symptom severity and depression (Chwastiak & Von Korff, 2003; Ertuğrul & Uluğ, 2002; Kim et al., 2005), its relation to objective measures (i.e., tests of functional capacity) is somewhat more variable. Cognition (i.e., performance on neuropsychological tests) is most often found to be related to measures of functional capacity, rather than reported disability (Ertuğrul & Uluğ, 2002; McKibbin et al., 2004; Rempfer, Hamera, Brown, & Cromwell, 2003; Twamley et al., 2002), although this latter relationship does have empirical support (M. F. Green, 1996; Kim et al., 2005). Discrepancies in self-report and objective measures of capacity found in studies of patients with schizophrenia, in particular, may be somewhat accounted for by impaired awareness in this population (Doyle et al., 1999; Harvey, 2010).

As WHODAS 2.0 has gained empirical support, the American Psychiatric Association (APA) has recommended that with the publication of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), the measure be used as a replacement for the Global Assessment of Functioning (GAF) from previous editions of the DSM (e.g., DSM-IV-TR; American Psychiatric Association, 2000). Along with this recommendation is the implication that WHODAS 2.0 be used in the context of both medical and legal disability evaluations. Gold (2014) has criticized APA's outright adoption of WHODAS 2.0 on the grounds that the instrument lacks embedded indices of response validity. That is, there are no internal measures to detect response bias (e.g., infrequently endorsed items or empirically-derived cutoff scores, above which scores are believed to be reflect exaggerated responding). An association between symptom validity performance and self-reported disability has been established previously by Kaye and colleagues (2014), who reported levels of disability and endorsement of symptom validity items were positively and significantly correlated. This raises concern that WHODAS 2.0 would likewise be influenced by the exaggeration of symptoms. Importantly, the present study draws a distinction between *symptom validity tests* (SVTs) and *performance validity tests* (PVTs), as clearly stated by Larrabee (2012). The former refers exclusively to measures of over-endorsement of symptoms, and the latter to measures designed to detect suboptimal effort on neuropsychological tests. The importance of assessing validity of responding is also a practice guideline for both national boarding organizations in clinical neuropsychology, the American Association of Clinical Neuropsychology (AACN; Heilbronner et al., 2009) and the National Academy

of Neuropsychology (NAN; Bush et al., 2005). If WHODAS 2.0 is to be accepted as an adequate clinical tool to assess a patient's functional limitations as intended, the relative influence of symptom validity on WHODAS 2.0 item endorsement must be critically and carefully examined.

SVTs and PVTs are notably absent from most previous studies of WHODAS 2.0. In fact, no studies of the relationship of cognition and item endorsement on WHODAS 2.0 include reference to symptom validity, though one recent study found that poor PVT performance (i.e., suboptimal effort on cognitive measures) was shown to be associated with higher WHODAS 2.0 scores in a sample of Veterans of recent conflicts (Clark, Amick, Fortier, Milberg, & McGlinchey, 2014). Although PVTs are demonstrably distinct from SVTs, this finding supports the notion that the relationship of cognition and disability may be more nuanced than current literature suggests. Specifically, to understand how cognition may be related to self-reported disability, awareness of the influence that symptom validity has on this relationship is vital. Considering these constructs using a robust and comprehensive statistical approach is a unique contribution of the present study.

Individuals undergoing evaluation for seizure disorders provide a unique opportunity to examine the intersection of concerns related to cognition, symptom validity, and disability. Cognition and disability have been well-characterized in these populations. While many of these patients are confirmed to have epilepsy, a large number (Bodde et al., 2009) are diagnosed with psychogenic nonepileptic events (PNEE). These are episodes which may resemble epileptic seizures behaviorally, but without accompanying

neuronal discharge (Binder & Salinsky, 2007). Cognitive complaints are commonly reported by these patients (Binder, Kindermann, Heaton, & Salinsky, 1998; Hermann & Seidenberg, 2007; Lee, 2010; Matsuoka, 2001; Willment, Hill, Baslet, & Loring, 2015), as are functional limitations, including disablement from employment (Breier, Fuchs, & Brookshire, 1998; Clarke, Upton, & Castellanos, 2006; Gilliam, Hecimovic, & Sheline, 2003; Krawetz et al., 2001; Walczak et al., 1995). To a lesser extent, symptom and performance validities have also been investigated in this population. Most of these studies have found SVTs contribute useful diagnostic information, such that patients with PNEE are more likely to fail score above cutoffs, and endorse unusual or medically unexplainable symptoms compared to those with ES (Benge et al., 2012; Locke et al., 2010; D. J. Williamson, Drane, & Stroup, 2007). Further, patients with PNEE may be more likely to fail PVTs than patients with ES (Drane et al., 2006; D. Williamson et al., 2004; D. J. Williamson et al., 2007; but also see Dodrill, 2008).

Measuring Disability Using WHODAS 2.0.

In 1988, WHO published the Psychiatric Disability Assessment Schedule (WHO/DAS; World Health Organization, 1988). The original measure's intent was to standardize the assessment of social functioning of patients with mental health disorders, separate from the clinical characteristics of the disorders themselves. This measure provided clinical information about a patient's functioning that was not adequately addressed in the International Classification of Impairments, Disabilities, and Handicaps (ICIDH; World Health Organization, 1980). More recently, WHO published an updated and revised model of classification of health and functioning, in the form of the ICF (World Health

Organization, 2001). This revised classification system supported a generic view of disability and warranted the development of a conceptually different, standardized, cross-culturally adaptable measure of disability (Üstün et al., 2010). The full history of the WHO/DAS and development of WHODAS 2.0 is described elsewhere (Federici & Meloni, 2010; Federici, Meloni, & Presti, 2009; Üstün, 2010; Üstün et al., 2010).

Clinically, WHODAS 2.0, in its full 36-item form or abbreviated 12-item form, has been validated as a measure of disability in the ICF framework across a wide range of medical and mental health conditions and settings, including arthritis (Baron et al., 2008); mild cognitive impairment (Bombin et al., 2012); Huntington's Disease (Carlozzi et al., 2015); hearing loss (Chisolm, Abrams, McArdle, Wilson, & Doyle, 2005); mood, anxiety, psychotic, and chronic pain disorders (Chopra, Couper, & Herrman, 2004; Chwastiak & Von Korff, 2003; Guilera et al., 2015; Janca et al., 1996; Konecky, Meyer, Marx, Kimbrel, & Morissette, 2014; McKibbin et al., 2004; Perini, Slade, & Andrews, 2006); stroke and rehabilitation (Küçükdeveci et al., 2013; Kulnik & Nikoletou, 2013; Pösl et al., 2007); and others (Federici, Meloni, & Presti, 2009; Gallagher & Mulvany, 2004; Garin et al., 2010; Hudson et al., 2008). The above studies have consistently reported excellent reliability and validity of WHODAS 2.0, providing support for its conceptual development and utility as a clinical tool.

Üstün and colleagues (2010) describe the findings from a series of studies conducted across the globe between 2000 and 2010. The studies were conducted in healthy populations, as well as in samples of patients with psychiatric and medical illnesses. WHODAS 2.0 has been translated into 31 languages and cross-culturally

validated. These extensive studies used Confirmatory Factor Analysis to confirm a stable two-level factor structure. The first level consists of a general disability factor, and the second level consists of the six WHODAS 2.0 subdomains; cognition (in text, "understanding/communication"), mobility, self-care, getting along, life activities (including household and work-related items), and social participation. Cronbach's alpha coefficient for the understanding/communication domain was 0.86; for mobility, 0.90; for self-care, 0.79; for getting along, 0.84; for home life activities, 0.98; for work/school life activities, 0.96; and for participation in society, 0.84. Cronbach's alpha for the entire measure was 0.96. Test-retest reliability was good, ranging from $r = 0.93$ to $r = 0.96$ within domains, and $r = 0.98$ overall. WHODAS 2.0 has demonstrated concurrent validity with a number of other quality of life and functioning measures, particularly when correlations between domains measuring similar constructs were computed. For example, the mobility domains of WHODAS 2.0 and the Functional Independence Measure (FIM; Granger, Hamilton, Linacre, Heinemann, & Wright, 1993) demonstrated stronger concurrent validity ($r = -0.78$) than the WHODAS 2.0 interpersonal domain and FIM overall score ($r = -0.34$).The overall WHODAS 2.0 score was highly correlated with overall scores on the FIM ($r = 0.68$), London Handicap Scale ($r = 0.75$; Harwood & Ebrahim, 1995), and WHO Quality of Life scale ($r = 0.68$; WHOQOL Group, 1993). Analysis of means in WHODAS 2.0 total and domain scores revealed differences in domain scores between clinical subgroups of patients (Üstün et al., 2010). These differences were in the expected direction based on the clinical subgroup (e.g., the group with physical health problems reported greater difficulty in mobility than the group with

mental health problems). Additional differences included the following: groups with mental health and substance use problems showed greater disability in the interpersonal domain; groups with mental health, alcohol, and substance use problems reported greater difficulty in the cognitive domain than the physical health problems group; and the general population reported less disability across all domains and total score, as compared to other clinical subgroups.

Measuring disability in the epilepsy monitoring unit (EMU).

Commonly used measures of disability for patients undergoing video EEG monitoring include the Quality of Life in Epilepsy Inventory (QOLIE; Breier et al., 1998; Cramer et al., 1998; Grudzinski, Hakim, Coons, & Labiner, 1998; Harden et al., 2007), and SF-36 (Al Marzooqi, Baker, Reilly, & Salmon, 2004; Lawton, Mayor, Howlett, & Reuber, 2009; Zeber, Copeland, Amuan, Cramer, & Pugh, 2007). Patients with epilepsy are likely to experience functional limitations and report increased disability as compared to healthy controls (Breier et al., 1998; Francis & Baker, 1999; Gilliam et al., 2003; Harden et al., 2007). Interestingly, patients with PNEE may be more likely than those with ES to report greater disability (Lawton et al., 2009), report significantly more psychiatric concerns (Asmussen, Kirlin, Gale, & Chung, 2009; Dworetzky et al., 2005), and seek financial benefits (Binder, Salinsky, & Smith, 1994), despite vastly different etiologies. Epilepsy and PNEE have also both been associated with decreased ability to maintain employment (Clarke et al., 2006; Smeets, van Lierop, Vanhoutvin, Aldenkamp, & Nijhuis, 2007). High rates of psychiatric comorbidity common in ES and PNEE are also associated with decrements in quality of life (Zeber et al., 2007), though in patients with

confirmed epilepsy, this diagnosis alone independently accounts for a large proportion of variance in quality of life (Kessler et al., 2012). As WHODAS 2.0 is not specific to etiology of disability, the measure could reliably be used to gather data on functional limitations in patients in the EMU.

However, surprisingly few studies have examined WHODAS 2.0 in the context of patients undergoing evaluation of seizure disorders. Of these, one noted significantly elevated WHODAS 2.0 in patients with epilepsy relative to healthy controls (Kessler et al., 2012), and another found WHODAS 2.0 to be sensitive to change in patients with mesial temporal lobe epilepsy who underwent anterior temporal lobectomy (Cankurtaran, Ulug, Saygi, Tiryaki, & Akalan, 2005). A large-scale study to generate a registry to assess the social and economic impact of epilepsy has been initiated in Australia (Hackett, Glozier, Martiniuk, Jan, & Anderson, 2011), which will include WHODAS 2.0 as a measure of psychosocial functioning, however findings from this study have not yet been published. No publications to date have investigated WHODAS 2.0 in patients with PNEE.

Cognitive performance in the EMU.

Cognitive complaints are common in patients being evaluated for seizure disorders. Individuals diagnosed with epilepsy have shown relatively worse neurocognitive test performance compared to healthy controls on tasks of visual motor skills, mental flexibility, psychomotor processing speed, sustained attention, visual and verbal memory, language, and even full scale IQ (Äikiä, Salmenperä, Partanen, & Kälviäinen, 2001; Jokeit & Ebner, 2002; Mayeux, Brandt, Rosen, & Benson, 1980; Ogunrin, Adamolekun,

Ogunniyi, & Aldenkamp, 2000; Pulliainen, Kuikka, & Jokelainen, 2000; Roeschl-Heils, Bledowski, Elger, Heils, & Helmstaedter, 2002). Similarly, PNEE has been associated with performance on neuropsychological measures which may be lower than expected. Wilkus and colleagues (1984) reported that the proportion of neuropsychological test scores which were 'abnormal' (i.e., lower than expected) was comparable in ES and PNEE. Comparable performance on neuropsychological tests between patients with ES and PNEE has been reported in other studies as well (Binder et al., 1998; M. Brown, Levin, Ramsay, Katz, & Duchowny, 1991). However, given the relatively high rate of PVT failure noted in patients with PNEE (Binder et al., 1998), it is not surprising that after controlling for suboptimal effort, patients with PNEE showed less objective evidence of cognitive impairment compared to those with ES (Drane et al., 2006). The prevalence of subjective and objective cognitive impairment in these populations underscore the need for thorough neuropsychological evaluation of these patients.

Symptom Validity in the EMU.

Medical and psychiatric evaluations are generally reported to see lower rates of malingering and symptom exaggeration relative to other medical disability samples (e.g., TBI litigants; Mittenberg, Patton, Canyock, & Condit, 2002). However, a diagnosis of epilepsy is frequently associated with disability, and therefore provides an opportunity for secondary gain (that is, external financial compensation). PVTs can therefore still be a useful tool for differentiating neurologic versus psychiatric causes of cognitive dysfunction (D. Williamson et al., 2004; D. J. Williamson et al., 2007). PVTs have demonstrated diagnostic utility in the EMU, such that patients with PNEE are more likely

to fail PVTs (Drane et al., 2006). Notably, this study also showed that very few patients with epilepsy (8%) failed PVT compared to patients with PNEE (49%). After controlling for PVT performance, these authors showed that patients with PNEE performed significantly better on a neuropsychological test battery compared to their counterparts with ES. Drane et al. also showed that symptom reporting varies between those with ES and PNEE. The latter group more frequently reported "unverifiable" symptoms including fibromyalgia and chronic pain. Benge (2012) similarly found significantly elevated scores on a measure of bizarre symptoms, the Structured Inventory of Malingered Symptomatology (SIMS) in PNEE compared to ES. In broader context, the infrequent somatic complaints (Fs) scale of MMPI-2-RF was shown to be sensitive to somatic malingering (Sellbom, Wygant, & Bagby, 2012). Further, Fs was not only significantly elevated in somatic malingering, but in confirmed somatoform disorders, as compared to other general medical conditions. Similarly, Locke and colleagues (2010) found that patients with PNEE obtained significantly higher scores on the Fs scale of MMPI-2-RF, though in their sample it did not contribute diagnostic information.

Symptom Validity and Disability.

In a number of samples of psychiatric samples, self-reported functional limitation is predicted by symptom severity (Bowie et al., 2010; Ertuğrul & Uluğ, 2002; McKibbin et al., 2004). It stands to reason that in the context of exaggerated symptom reporting (i.e., SVT failure), that report of disability would be similarly exaggerated, although this has not been extensively studied. Kaye (2014) reported such a finding in a sample of female Veterans with post-traumatic stress disorder (PTSD). In fact, SVT performance was not

only highly correlated with self-reported disability, PTSD symptom severity, and depression, but was the only significant predictor of disability.

As SVT and PVT are frequently used interchangeably, the majority of similar studies examine self-reported disability in the context of PVT performance. While not unrelated, these findings cannot be used to infer the relationship with SVT performance, as previously discussed. To date, WHODAS 2.0 has not been evaluated in the context of SVT, and specific concerns about biased responding have not been addressed in the literature. Clark and colleagues (2014) found significant differences in symptom rating scales between those who passed versus failed PVTs. Of particular interest, these authors also reported significantly higher WHODAS 2.0 scores in those failing PVTs, though no additional analysis was conducted using WHODAS 2.0. PVT performance has been demonstrated to be predictive of self-reported community reintegration in a sample of Veterans with a history of mild traumatic brain injury (Lippa et al., 2014). These findings underscore the importance of extending study into disability and SVT performance.

Aims of the current study

Aim 1. The primary aim of the study is to examine the independent relations among cognition, symptom validity, depression, and self-reported disability, with a particular emphasis on the extent to which symptom validity influences the relationship of cognition and self-reported disability. Through structural equation modeling (SEM), the strength of independent predictive relationships and their interaction will be explored in a robust statistical framework. Previous studies have addressed only the role of depression in this relationship, not investigating the influence of symptom validity, which

is a significant concern in a number of patient populations, particularly those under evaluation for medical or psychiatric disability and compensation. Additional exploratory analyses will consider the relative influence of performance validity, to further clarify the nature of differences in relationships.

Aim 2. A secondary aim of the study is to examine differences in WHODAS 2.0 scores between participants who score above and below recommended cutoffs on SVT. The unique contribution of the present study is in examining this in a population which tends to exhibit a relatively high rate of symptom exaggeration, thereby improving robustness of findings. Exploratory analyses will expand understanding of these relationships by incorporating performance validity considerations.

Hypotheses

Aim 1. An investigation of independent relationships among variables of interest was expected to yield a significant and moderately strong relationship between symptom validity endorsement and WHODAS 2.0. Consistent with previous literature, the correlation of WHODAS 2.0 with cognition, operationalized by an overall test battery mean, was expected to be weak. Similarly, cognition was expected to correlate weakly to moderately with symptom validity. Indicators of depression were expected to correlate significantly with both symptom validity and disability, and be moderately strong. Structural equation modeling was used to further assess these relationships, with particular interest in whether the relationship between cognition and disability is moderated by symptom validity. Specifically, more ‘valid’ symptom endorsement will be associated with a stronger relationship between cognition and disability.

Aim 2. Patients who exceed the recommended SVT cutoff were expected to report significantly higher functional limitation on WHODAS 2.0 total score as compared to patients whose symptom endorsement fell in the valid range. A moderate effect size was expected. Post hoc analyses will examine group differences within WHODAS 2.0 domains, though specific predictions regarding domains were not made *a priori*.

Methods

Participants

The present study retrospectively examined data from Veterans who were admitted to the epilepsy monitoring unit (EMU) of the Michael E. DeBakey VA Medical Center (MEDVAMC) between January 2015 and December 2016. Patients with documented dementia (moderate to severe) or schizophrenia were excluded from the study, as these conditions may impact comprehension and/or completion of self-report measures. Participants were 285 Veterans (20.7% female) over the age of 18 (M = 50.20, SD = 14.23) residing within the catchment area of the MEDVAMC. One hundred ninety participants (66.9%) were non-Hispanic Caucasian, 76 (26.8%) were African-American, 15 (5.3%) were Hispanic, one (0.4%) identified as multiracial, and two (0.7%) identified as ‘other.’ The current study has obtained approval from the Baylor College of Medicine IRB (Protocol Number: H-39612) and MEDVAMC Research and Development Committee (ID Number: 16L02.H), and has been approved for reliance on an external IRB (Baylor College of Medicine) by the University of Houston Committee for the Protection of Human Subjects (study ID: 00000179).

Typical length of stay on the EMU was five days, during which time the Veterans are monitored by video-EEG (VEEG) for paroxysmal events. Over the course of the week, Veterans also undergo extensive evaluation, including obtaining comprehensive psychosocial and medical histories, neuroimaging, and a brief neuropsychological evaluation which includes a number of measures of cognition, mood, performance and symptom validity, functioning, quality of life, and personality. Diagnosis is reached using findings from all aforementioned procedures. The current sample includes patients diagnosed with VEEG-confirmed or probable epilepsy (30.1%), PNEE (37.5%), mixed (the presence of both epileptic seizures and PNEE; 4.6%), or other diagnoses (including, but not limited to, syncopal episodes, panic disorder, anxiety disorders, and misinterpretation of benign somatic symptoms; 14.7%). Definitive diagnosis could not be reached for 12.3% of patients.

Measures

Neuropsychological evaluations were conducted with each Veteran during their admission to the epilepsy monitoring unit. This includes cognitive, emotional, personality, disability, and performance and symptom validity assessment. Cognitive measures include common neuropsychological tests. Self-report versions of all emotional and disability inventories were used. All assessment instruments relevant to the current study are described in detail below.

Disability/Functioning. WHODAS 2.0 (Appendix 1; Üstün, 2010) requires respondents to rate, on a scale of 1 (none) to 5 (extreme/severe), how much difficulty they have experienced engaging in a variety of tasks in the 30 days preceding the

administration of the measure. WHODAS 2.0 captures perceived disability across six life domains: understanding and communication, mobility, self-care, getting along, life activities (household and work), and social participation. Individual items were developed from ICF items related to the ICF 'Activity and Participation' component (World Health Organization, 2001). WHODAS 2.0 items are independent of etiology of disability, allowing the measure to be applied broadly across medical and mental health conditions.

Scores on WHODAS 2.0 can be computed according to 'simple' or 'complex' scoring procedures. According to simple scoring, item responses are summed and provide an "overall' disability score with a range of 0 – 144 (sometimes represented as a percentage of the maximum), in addition to mean domain scores, with higher scores indicating more severe disability. The present study will utilize scores determined by the complex scoring protocol, which applies differential weights to individual items, developed based on item response theory (IRT), yielding standardized overall and domain scores ranging from 0-100. Again, increasing scores denote increasing disability. WHO provides SPSS syntax that will be used to perform this transformation (Üstün, 2010). This scoring approach was established to enhance applicability of WHODAS 2.0 scores across diagnostic and cultural populations (Üstün et al., 2010), and although most WHODAS 2.0 literature has used complex scoring, differential item weights have not been published for specific populations. WHO has recommended against interpreting results of simple scoring, and has not published normative data based on this approach.

Simple scoring, however, may be preferable in busy clinical settings where an overall measure of disability conveys adequate information.

Within the eight items of the "life activities" domain of WHODAS 2.0 are four optional items specific to remunerated work. Given that many patients in the current sample are unemployed, for a variety of reasons including disablement from employment, a large number of respondents did not endorse these items. According to WHODAS 2.0 scoring protocol (Üstün, 2010), following the complex scoring procedure for the remaining 32 items yields a total disability score that is comparable to that obtained from the full 36 item version. For this reason, items 5-8 from the "life activities" domain were excluded from analyses. Thus, the four non-work ("household activities") items represent the "life activities" domain.

In the context of the measurement and structural models of interest to the current study, disability was represented as a latent variable, with six WHODAS 2.0 subdomains as indicators. This model is supported by prior factor analytic studies which have consistently demonstrated the six-factor structure of WHODAS 2.0 across a range of populations (Buist-Bouwman et al., 2008; Federici, Meloni, Mancini, Lauriola, & Olivetti Belardinelli, 2009; Federici, Meloni, & Presti, 2009; Pösl et al., 2007; Üstün et al., 2010).

Depression. The Beck Depression Inventory, second edition (BDI-II; Beck, Steer, & Brown, 1996) is a 21-item self-report inventory which asks the respondent to rate the intensity of symptoms of depression over the preceding two weeks. Each item can be answered on a 0 to 3 scale, for a total score range of 0 to 63, with higher scores indicating more severe depression. The BDI-II has been shown to be a reliable and valid measure of

the cognitive/affective as well as somatic components of depression (Storch, Roberti, & Roth, 2004).

The Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001) is composed of nine items from the self-report version of the depression module of the Primary Care Evaluation of Mental Disorders (PRIME-MD; Spitzer et al., 1994), a diagnostic instrument for assessing psychiatric disorders. Item responses range from 0 to 3, for a total score range of 0 to 27, with higher scores indicating greater frequency and intensity of depressive symptoms. The PHQ-9 has been established to have good convergent validity with other measures of depression (Martin, Rief, Klaiberg, & Braehler, 2006).

The Minnesota Multiphasic Personality Inventory – 2 – Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) is a 338-item self-report instrument for the assessment of psychopathology and personality traits. The MMPI-2-RF is a shortened, restructured form of the MMPI-2 which has enhanced psychometric properties, and involved the development of non-overlapping restructured clinical (RC) scales. These scales have demonstrated good discriminant validity (Sellbom, Bagby, Kushner, Quilty, & Ayearst, 2011). In particular, these authors found RC2 (Low Positive Emotions) elevations to be associated with a higher probability of a diagnosis of major depression. This is not surprising given that low positive emotionality is, in particular, a defining characteristic of depression (Brown, Chorpita, & Barlow, 1998; Sellbom, Ben-Porath, & Bagby, 2008).

In the measurement and structural models of relevance to this study, depression was represented as a latent variable, with BDI-II, PHQ-9, and RC2 total scores as indicators. Consistent with previously published literature, this latent variable acted as a covariate of disability in the measurement model. The structural model included this latent variable of depression as a predictor of disability.

Symptom Validity. The MMPI-2-RF (Ben-Porath & Tellegen, 2008) contains a number of validity scales designed to detect non-credible responding. Of interest in the current study is Fs (infrequent somatic responses). This scale is new to the Restructured Form of MMPI-2, and includes 16 infrequently-endorsed items. For the present study, item endorsement on Fs was analyzed and reported in terms of normative *t*-score conversions, as is generally consistent with similar literature. The Fs scale has been shown to differentiate credible versus non-credible symptom reporting across a number of samples (Locke et al., 2010; Sellbom et al., 2012; Wygant et al., 2009), and has a relatively lower rate of false positive errors as compared to other symptom validity scales such as FBS-r (Butcher, Arbisi, Atlis, & McNulty, 2008). In the measurement and structural models, symptom validity scores were treated as continuous, though for means comparisons, were dichotomized into "valid" and "invalid" ranges. Invalid symptom reporting was operationalized as a score at or above 100*T* on Fs, as recommended in the MMPI-2-RF Administrative Manual (Ben-Porath & Tellegen, 2008). Scores below this cutoff were considered valid.

Cognition. Participants were administered a brief neurocognitive test battery assessing the domains of learning and memory, processing speed, attention, and

executive functioning, as described below. Raw scores on each neuropsychological test administered were converted to a T-score (M = 50, SD = 10) based on the available normative references. The mean of all four domain scores yielded a global neuropsychological test performance variable similar to that used in previously published literature (Bombin et al., 2012), an overall test battery mean (OTBM). This variable represented "cognition" in subsequent analyses.

Learning and memory. Participants were administered the Logical Memory subtest of the Wechsler Memory Scale, 4th edition (WMS-IV; Wechsler, 2009). During the first trial (LMI), the examinee is read a short story, and asked to immediately repeat the story in as great of detail as possible. Then the examinee is read a second short story, and the task is repeated. Following a delay of between 20-30 minutes, and without being forewarned, the second trial (LM II) is administered, and the examinee is asked to recall as much information from each of the stories as possible. Each of these trials yields an age-normed scaled score, based on data from the validation sample. Each scaled score was converted to a T-score, and the mean of the two represented performance in the domain of *learning and memory*.

Processing speed. Patients were administered two subtests (Symbol Search and Coding) of the Wechsler Adult Intelligence Scale, 4th edition (WAIS-IV; Wechsler, 2008). Symbol Search is a timed measure of the examinee's ability to identify a target shape among distractors in a visual array. The examinee has two minutes to complete as much of the task as possible. The measure relies not only on processing speed, but also short-term visual memory, attention to visual detail, and visual discrimination. The

Coding subtest is another timed measure, requiring the examinee to match symbols to their corresponding number according to a digit-symbol code that is provided. Again, the examinee has two minutes to complete this task. In addition to processing speed, the Coding subtest engages fine motor dexterity, associative learning, and cognitive flexibility. Each of these subtests yields an age-normed scaled score based on the validation sample. From the sum of these scaled scores, a Processing Speed Index (PSI) is derived, a composite metric of cognitive processing speed. For this study, this standard score was then converted to a T-score. The PSI is used instead of the individual subtest age-normed scores to somewhat control for non-processing speed factors that are unshared between the tests, as described above.

Examinees also complete the Trail Making Test (TMT; Reitan, 1958), a timed visual scanning and sequencing measure of processing speed. Trial A of the task (TMT-A) has examinees rapidly sequence numbers 1 through 25. The time to completion is converted to a demographically-normed T-score based on widely available normative standards (Heaton, Grant, & Matthews, 1991). Performance in the *processing speed* domain was defined by the mean of WAIS-IV PSI and TMT-A T-scores.

Attention. Patients were administered the Digit Span (DS) subtest of the WAIS-IV (Wechsler, 2008). This subtest of composed of three units, the first of which requires the examinee to repeat an increasingly long series of numbers. The examinee is then asked to reverse the given number series. Finally, the examinee is given series of numbers which he or she must recite in order from the lowest to the highest number. All three scores are

summed and the resulting total yields an age-normed scaled score based on the validation sample. This scaled score was then converted to a T-score to reflect the *attention* domain.

Executive functioning. Examinees completed trial B of the TMT (TMT-B), a timed test requiring the examinee to visually scan and sequence alternating numbers and letters (i.e., 1-A-2-B…13). As with trial A of TMT, the time of completion yields a demographically-normed T-score (Heaton et al., 1991). While the test is largely dependent on graphomotor processing, it is commonly used as a metric for executive functions (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Chaytor, Schmitter-Edgecombe, & Burr, 2006; Lezak, 1995) due to the set-shifting and cognitive flexibility demands of the task.

Performance validity. Patients were administered Green's Word Memory Test (WMT; P. Green, 2005), a common recognition memory-based task of performance validity. The computer-based measure consists of 20 word pairs (40 words total) presented twice visually, followed immediately by a forced-choice recognition memory test. Throughout the test, subjects are provided with information about the accuracy of their responses, as indicated by green or red highlighting of their responses. Following an approximately 30-minute delay, the same forced-choice recognition test is administered again. The scoring program's output includes the percentage correct on each trial, and a "consistency" index that is the percentage concordance of the subject's responses. If the score on any one of these primary subtests falls at or below 82.5%, performance is considered "invalid." Importantly with regard to the population currently under study, "WMT primary effort subtests are generally insensitive to known temporal lobe

pathology," (Eichstaedt et al., 2014, p. 941). This suggests that WMT failure is sensitive and specific to detecting suboptimal performance in this population. For analysis, participants were dichotomized based on passing or failing the WMT.

Analysis

Analyses of correlations, mean comparisons, and measurement and structural models, were conducted using Statistical Package for the Social Sciences, version 22 (SPSS 22.0; IBM, Inc., 2013), and the open-source statistical program R (R Core Team, 2017) and the associated *lavaan* SEM package (Rosseel, 2012). Analysis of independent relationships among indicators of interest (OTBM, SVT performance, indicators of depression, and WHODAS 2.0), correlations and their respective 95% confidence intervals were assessed. Figure 1 shows the proposed measurement model. The model includes two factors; disability, indicated by the six WHODAS 2.0 subdomains, and depression, indicated by two self-report measures of depression as described below. The model was generated *a priori* based on established theoretical relationship between these two factors (Guilera et al., 2015; Kim et al., 2005), and validity of the indicators of both depression (Martin et al., 2006; Storch et al., 2004) and disability (Üstün et al., 2010). Confirmatory Factor Analysis (CFA) approaches were used to test overall fit of the measurement, which was expected to be good.

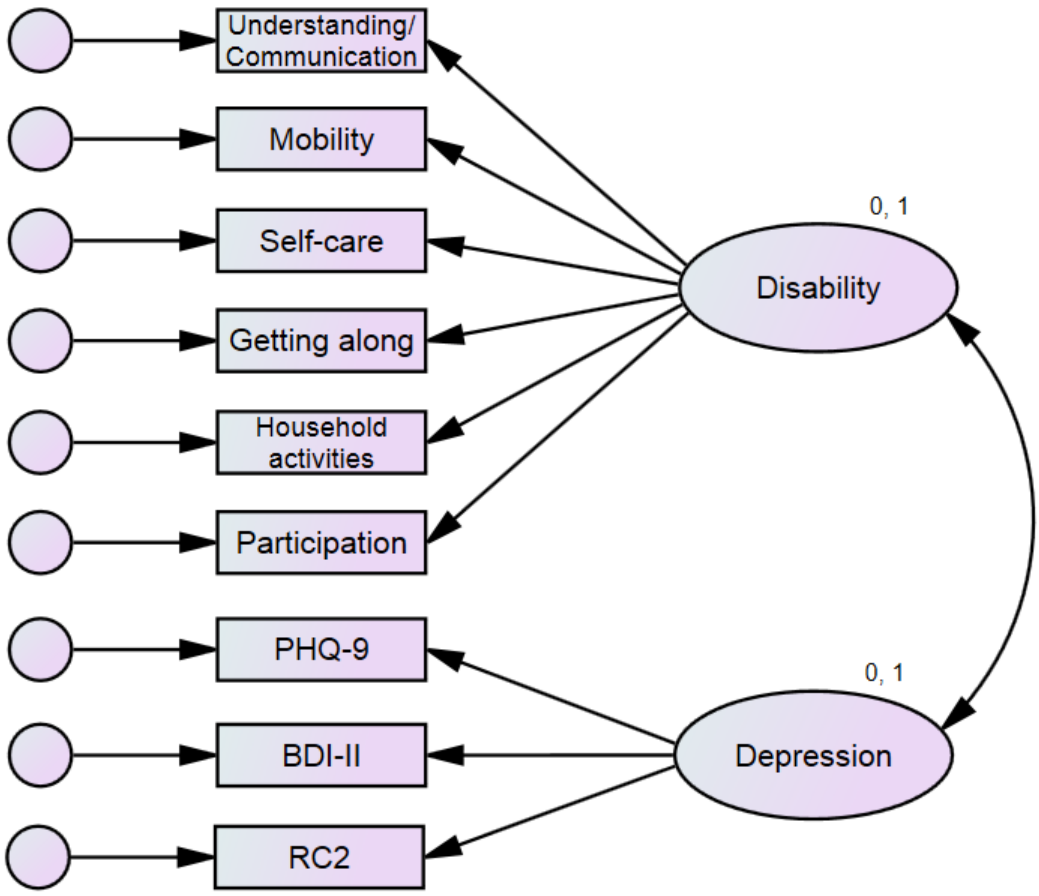

Figure 1. Proposed Measurement Model. The variances of the latent variables, Disability and Depression, were set to 1. The six Disability indicators correspond to WHODAS 2.0 subdomains. PHQ-9, Patient Health Questionniare – 9 Item; BDI-II, Beck Depression Inventory, 2nd edition; RC2, Minnesota Multiphasic Personality Inventory, 2nd edition, Restructured Form, Restructure Clinical Scale 2 (Low Positive Emotion).

Model fit was assessed with both relative and absolute indices of model fit. Relative indices compare chi square for a null model in which all variables are uncorrelated to the hypothesized model, and include the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), and the sample-size adjusted comparative fit index (CFI; Bentler, 1990). Recommended cutoffs for TLI and CFI are 0.95, with values closer to 1 indicating better fit (Hu & Bentler, 1999). Absolute indices of model fit test how well the hypothesized model fit the sample data, and commonly include chi square (Fox, 2010),

root mean square error of the approximation (RMSEA; Steiger & Lind, 1980), and standardized root mean square residual (SRMR; Schreiber, Nora, Stage, Barlow, & King, 2006). Failure to reject the null hypothesis of the chi square test indicates good model fit, although the test is likely to produce type I errors in the presence of large samples, strong correlations among variables, and non-normality of variables (Kenny, 2012). RMSEA analyzes the discrepancy between a hypothesized model with optimal parameter estimates and the population covariance matrix. The value of RMSEA is reported in addition to its 90% confidence interval, to gauge the precision of the estimate, and likelihood of acceptable model fit. While there is some variability in the literature, values should fall below 0.10 for acceptable fit, while values below 0.08 indicate good fit, with lower values indicating better fit (Browne, Cudeck, Bollen, & Long, 1993). An advantage of RMSEA is that it is not dependent on sample size. The discrepancy between the correlation matrix of the sample and hypothesized model is reported by SRMR. Model complexity does not adversely affect SRMR, and values below .08 are generally interpreted as reflecting acceptable model fit (Hu & Bentler, 1999).

Next, an expanded CFA model was generated and tested in the same manner, which included two additional relevant constructs, Cognition and Symptom Validity. These each had single indicators, the OTBM, and MMPI-2-RF Fs *t*-score, respectively. This measurement model is depicted in Figure 2.

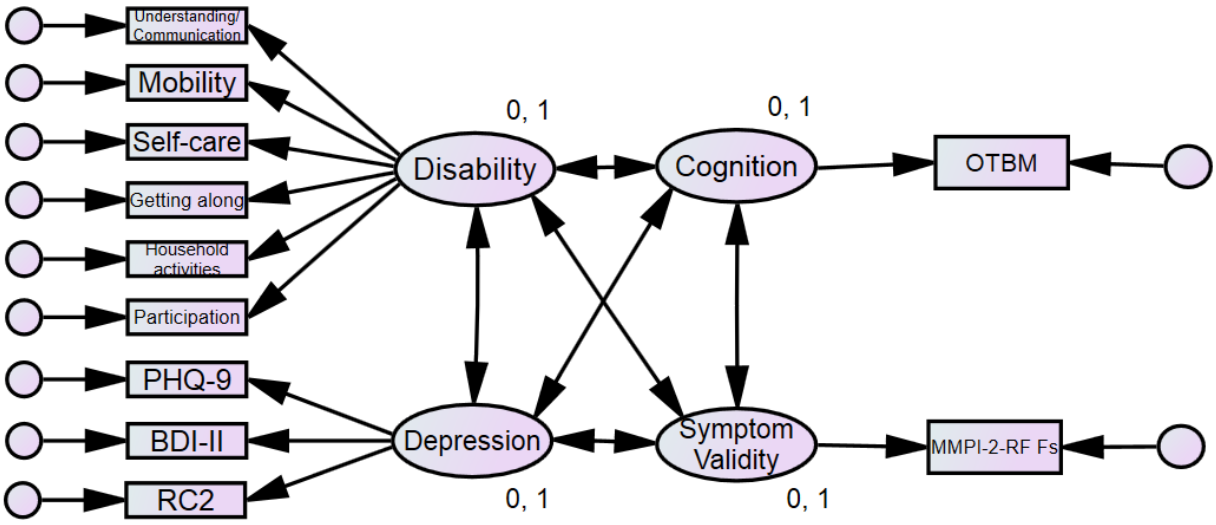

Figure 2. Expanded Measurement Model. All variances of latent constructs were set to 1. PHQ-9, Patient Health Questionnaire – 9 Item; BDI-II, Beck Depression Inventory, 2nd edition; RC2, Minnesota Multiphasic Personality Inventory, 2nd edition, Restructured Form, Restructure Clinical Scale 2 (Low Positive Emotion); OTBM, Overall Test Battery Mean; Fs, MMPI-2-RF Infrequent Somatic Complaints scale.

After confirming acceptable fit of the expanded measurement model, an initial structural model was assessed. This preliminary structural model included only the predictor main effects, not the interaction term. The aforementioned model fit indices (TLI, CFI, Chi square, RMSEA, SRMR) were investigated for structural model fit, and the standardized parameter estimates were inspected for strength and significance. In the interaction model, to test the moderation hypothesis, the product term of the standardized values of OTBM and Fs was included as an additional predictor. The presence of a significant, large parameter estimate for the interaction would suggest that a significant interaction effect exists in the prediction of self-reported disability.

Data were screened for univariate normality using Shapiro-Wilk's test. To screen for multivariate normality, SPSS macros developed by DeCarlo (1997) were used. These macros provide a number of statistics, most importantly Mardia's multivariate skew and

kurtosis (Mardia, 1980), and a plot of the squared Mahalanobis distances (Penny, 1996), among others.

Given the medical and psychiatric complexity of the sample investigated in the present study, missing data was anticipated. In addition to inspection of proportions of missing data, Little's test for data Missing Completely at Random (Little, 1988) was conducted. Missing data was handled by full information maximum likelihood estimation (FIML; Hartley & Hocking, 1971), for relevant analyses, which is likely to produce unbiased parameter estimates in SEM (Dong & Peng, 2013; Enders & Bandalos, 2001).

Results

Data properties and sample characteristics

Statistical power. A number of general and somewhat discrepant "rules of thumb" for minimum sample size requirements for confirmatory factor analysis and structural equation modeling are pervasive and often critiqued in the literature (Fox, 2010; MacCallum, Widaman, Zhang, & Hong, 1999; Shieh, 2010). These range from simple cutoffs, (e.g., $N \geq 200; 400$) to number of subjects per estimated parameter (e.g., $N = 10$ or 20 subjects per parameter). However, these crude methods do not account for model complexity, factor loadings, number of indicators, etc. More sophisticated techniques for power analysis and minimum sample size calculations have been developed using simulation studies (MacCallum, Browne, & Sugawara, 1996; Muthén & Muthén, 2002; Wolf, Harrington, Clark, & Miller, 2013). It is also possible to use these methods to determine statistical power and minimum sample sizes for specific fit indices. One such method uses the statistical program *R* (R Core Team, 2017) and the *lavaan*

(Rosseel, 2012) package to calculate power for RMSEA (Preacher & Coffman, 2006). When applied to the current study's confirmatory factor analysis, there is an increased likelihood of failing to detect a small discrepancy between the null "optimally estimated" model and the alternative model (N = 282, df = 25, H_a RMSEA = 0.05, power = 65%). Using the same basic calculation protocol, the minimum sample size required to achieve 80% power is N = 367. Determining minimum sample size for adequate statistical power based on the considerations of Wolf et al. (2013), the following are examined: the number of factors, the number of indicators per factor, and the hypothesized factor loadings.

Normality. Shapiro-Wilk tests of univariate normality revealed significant departures from normality in distributions of scores on BDI-II, PHQ-9, Fs, RC2, and each individual domain of WHODAS 2.0 (all $p < .01$). Skewness values ranged from -0.02 to 0.72 (*SE* range = [0.15, 0.16]). Values of kurtosis ranged from -0.98 to -.20 (*SE* range = [0.29, 0.31]). These small skewness and kurtosis values suggest minimal departures from normality, and are likely significant due to the relatively large sample size (Tabachnick & Fidell, 2012). Visual inspection of Normal Q-Q plots did not reveal notable departures from normality, suggesting Shapiro-Wilk tests may have been overly sensitive. Further, transformation conducted on the data in question generally did not yield distributions which yielded a non-significant Shapiro-Wilk test of normality. Given that these values of skewness and kurtosis are unlikely to produce biased parameter estimates (Tabachnick & Fidell, 2012), data transformations were not used in analysis, to ease interpretation of resulting parameter estimates. The other predictor and outcome variables of interest were

normally distributed according to Shapiro-Wilk tests. WHODAS 2.0 total scores were normally distributed with a skewness of 0.08 (*SE* = 0.15), and a kurtosis of -0.61 (*SE* = 0.30). The distribution of OTBM was also normal, with a skewness of 0.21 (*SE* = 0.15) and kurtosis of 0.08 (*SE* = 0.30).

Perhaps unsurprisingly, given the findings from tests of univariate normality, tests of multivariate normality generally suggested that the multivariate distributions are non-normal. Mardia's test of multivariate skew (Mardia, 1980) was significant ($p < .01$), as was Mardia's test of multivariate kurtosis ($p < .05$), suggestive multivariate non-normality. Analysis of Mahalanobis Distances (critical $F(11, 206) = 33.28$, all distances < 29.00) suggest that no multivariate outliers were present in the data. For this reason, no observations were manually removed from further analysis.

Several accommodations were made to address the apparent departure from normality in the data. Most importantly, preliminary descriptive statistics and model analyses used robust maximum likelihood estimation (MLR; Huber, 1967) for parameter estimation. This approach generates identical parameter estimates to the standard maximum likelihood estimation, but yields adjusted standard errors and model fit statistics that are robust to violations of multivariate normality. The MLR chi-square test statistic is scaled by a correction factor described by Yuan and Bentler (2000). Additionally, where applicable, nonparametric tests were used, including Mann-Whitney *U* Test (Mann & Whitney, 1947) for means comparisons in secondary aims.

Missing data. According to strict management of missing data, a WHODAS 2.0 domain score was only calculated if all items within the domain were answered. Two

hundred thirty-six (82.8%) patients answered every item of WHODAS 2.0. An additional 20 participants lacked at least one domain score due to missing a single item within one or more domains. The WHODAS 2.0 publication manual stats that mean substitution can be used to address missing data in the case of a single missing item within a domain, without exerting a large influence on the distribution of scores (Üstün, 2010). This procedure was undertaken in order to yield complete WHODAS 2.0 data for 256 (89.8%) patients. Examination of rates of missing data for these and other variables revealed that two hundred and eighteen patients (76.5%) were missing no data. The remaining sixty-seven cases were missing data for at least one variable. Overall rate of missing datapoints was 9.9%, across all relevant variables. When demographic variables were included in this inspection, 7.3% were missing. Percent missing data from each variable are included in Table 1. Importantly, Little's MCAR test was not significant ($\chi^2 = 93.59$, $df = 202$, $p = .99$), suggesting that data was missing completely at random. While this is an important assumption for handling missing data appropriately, a caveat is that it is not possible to unequivocally rule out that participants with missing data may differ according to some unmeasured characteristic. As previously mentioned, FIML was the method by which missing data were handled in model analysis using the *lavaan* statistical package. Some analyses conducted required the exclusion of individual cases that were missing data on all variables relevant to those individual analyses. As such, *N* may be less than 285, and these deviations are noted in the text.

Descriptive statistics. Descriptive statistics for the sample are described in Table 1. Means and standard deviations of model variables and their relevant components are

based on MLR estimation methods, and apply FIML procedures to account for missing data. The WHODAS 2.0 total score mean found in the present study (*M* = 38.36, *SD* = 19.17) is higher than published scores for other populations, including patients with depression, chronic pain, breast cancer, and general chronic physical health conditions. Many of these populations and values are presented in Appendix 2.

Table 1.

Sample Characteristics and Model Variable Descriptive Statistics

Variable	Latent construct	Mean	*SD*	% Missing
PHQ-9 Total	Depression	11.57	7.00	11.2
BDI-II Total	Depression	19.70	12.08	2.8
MMPI-2-RF RC2	Depression	63.97	14.49	15.1
Understanding/ Communication	Disability	34.04	20.78	9.1
Mobility	Disability	40.30	25.35	8.8
Self-care	Disability	22.81	22.58	8.8
Getting along	Disability	36.74	25.95	9.1
Household activities	Disability	44.32	29.42	9.1
Participation	Disability	46.91	22.94	10.2
WHODAS 2.0 Total		38.36	19.17	10.2
MMPI-2-RF Fs		80.84	21.91	15.1
Digit Span (T)		41.82	8.25	4.2
WAIS-IV PSI (T)		41.01	9.54	4.6
WMS-IV LM I (T)		43.31	10.99	5.6
WMS-IV LM II (T)		41.43	11.40	5.6
TMT-A (T)		44.01	11.36	4.6
TMT-B (T)		44.31	11.31	8.4
OTBM		42.90	7.24	9.5

Note: N = 282. PHQ-9, Patient Health Questionnaire - 9 item; BDI-II, Beck Depression Inventory, 2nd edition; MMPI-2-RF, Minnesota Multiphasic Personality Inventory, Second edition, Restructured Form; RC2, Restructured Clinical Scale 2 (Low Positive Emotion); WHODAS, World Health Organization Disability Assessment Schedule 2.0; Fs, Infrequent Somatic Complaints scale; PSI, WAIS-IV, Wechsler Adult Intelligence Scale, 4th edition; PSI, Processing Speed Index; WMS-IV, Wechsler Memory Scale, 4th edition; LM I & II, Logical Memory I & II; TMT-A&B, Trail Making Test, parts A & B; OTBM, Overall Test Battery Mean

Aim 1

Correlations. Independent relationships among all variables of interest were assessed by examining zero-order correlations and their respective confidence intervals and significance levels. Correlations were derived from covariances provided by *lavaan*, and reflect estimates based on MLR and missing data accounted for by FIML. The results are displayed in Table 2.

Table 2.

Correlations and 95% Confidence Intervals for CFA and SEM Analyses

	PHQ-9	BDI-II	MMPI-2-RF RC2	Under./ Comm.	Mobility	Self-care	Getting along	Household Activities	Part.	MMPI-2-RF Fs	OTBM	OTBM x Fs
PHQ-9	1											
BDI-II	.77** (.64, .90)	1										
MMPI-2-RF RC2	.54** (.41, .67)	.58** (.44, .71)	1									
Under./ Comm.	.63** (.50, .76)	.54** (.41, .67)	.43** (.29, .57)	1								
Mobility	.40** (.27, .53)	.39** (.26, .52)	.35** (.21, .48)	.45** (.32, .58)	1							
Self-care	.36** (.23, .49)	.36** (.23, .50)	.30** (.17, .43)	.50** (.37, .63)	.60** (.47, .72)	1						
Getting along	.56** (.43, .69)	.48** (.35, .61)	.45** (.32, .58)	.62** (.48, .75)	.47** (.34, .59)	.56** (.43, .68)	1					
Household Activities	.48** (.35, .60)	.41** (.29, .54)	.40** (.27, .53)	.59** (.46, .73)	.67** (.55, .80)	.67** (.54, .79)	.57** (.44, .70)	1				
Participation	.60** (.47, .72)	.53** (.40, .66)	.41** (.28, .55)	.59** (.46, .72)	.58** (.46, .70)	.59** (.47, .71)	.60** (.47, .72)	.68** (.56, .81)	1			
MMPI-2-RF Fs	.54** (.41, .68)	.45** (.31, .60)	.38** (.25, .51)	.46** (.31, .59)	.31** (.17, .44)	.30** (.15, .45)	.36** (.23, .48)	.35** (.21, .48)	.36** (.23, .50)	1		
OTBM	-.10 (-.22, .02)	-.08 (-.19, .04)	.00 (-.13, .13)	-.20** (-.32, -.08)	-.04 (-.16, .08)	-.12 (-.25, .01)	-.06 (-.18, .06)	-.09 (-.22, .03)	-.08 (-.19, .04)	-.11 (-.23, .02)	1	
OTBM x Fs	.28** (.16, .40)	.24** (.11, .37)	.25** (.11, .39)	.14* (.00, .27)	.18** (.06, .30)	.10 (-.05, .25)	.18** (.06, .31)	.16* (.02, .29)	.18** (.06, .29)	.56** (.42, .71)	.76** (.59, .92)	1

Note: $N = 282$. *$p < .05$; **$p < .01$.

CFA, Confirmatory Factor Analysis; SEM, Structural Equation Modeling; PHQ-9, Patient Health Questionnaire - 9 item; BDI-II, Beck Depression Inventory, 2nd edition; MMPI-2-RF, Minnesota Multiphasic Personality Inventory, Second edition, Restructured Form; RC2, Restructured Clinical Scale 2 (Low Positive Emotion); WHODAS, World Health Organization Disability Assessment Schedule 2.0; Fs, MMPI-2-RF Infrequent Somatic Complaints scale; OTBM, Overall Test Battery Mean; OTBM x Fs, Interaction of OTBM and Fs.

Indicators of depression were strongly correlated (*r* range = .54 to .76, all $p <$ .01). Intercorrelations of WHODAS 2.0 domains were generally moderate to strong, ranging from $r = .45$ to .68, all $p < .01$. Correlations among indicators of depression and indicators of disability were generally weaker, ranging from $r = .35$ to .63, but were all significant at $p < .01$. In addition, correlations between Fs and other model variables were somewhat weaker, but still generally moderate, ranging from $r = .38$ to .54 (all $p < .01$) with depression indicators, $r = .30$ to .46 (all $p < .01$) with disability indicators, and Fs was uncorrelated with OTBM ($r = -.11$, $p = .10$). Interestingly, Fs was most strongly correlated with scores on the Understanding/communication subdomain of WHODAS 2.0, and most weakly correlated with the Self-care subdomain. OTBM was significantly but weakly correlated with WHODAS Understanding/Communication domain ($r = -.20$, $p < .01$). The interaction of OTBM and Fs (OTBM x Fs) was significantly but rather weakly correlated with depression indicators ($r = .24$ to .28, $p < .01$), and was very weakly correlated with disability indicators. The correlation with the Understanding/ Communication WHODAS domain score was $r = .14$, $p < .05$; Mobility, $r = .18$, $p < .01$; Self-care, $r = .10$, $p = .19$; Getting along, $r = .18$, $p < .01$; Household activities, $r = .16$, $p < .05$; and Participation, $r = .18$, $p < .01$.

Measurement model. As described above, a two-factor model was hypothesized to be confirmed in the measurement model. Where applicable, statistics that were scaled (Yuan-Bentler adjustment) for robust maximum likelihood estimation are reported. Model fit indices essentially fell just within acceptable ranges, $\chi^2(26, N = 281) = 89.98$, $p < .05$, Yuan-Bentler correction factor = 1.14, TLI = 0.92, CFI = 0.94, RMSEA = .100

(90% CI = .08, .12), SRMR = .05. The measurement model and standardized parameter estimates are presented in Figure 3. Table 3 shows standardized and unstandardized coefficients for each of the indicators. Variance accounted for by the constructs was moderately high. The lowest estimates were noted for the mobility domain (52% of variance accounted for by disability), and scores on MMPI-2-RF RC2 (42% of variance accounted for by the construct of depression). These estimates are slightly lower than those reported in the WHODAS 2.0 formative studies (Üstün et al., 2010).

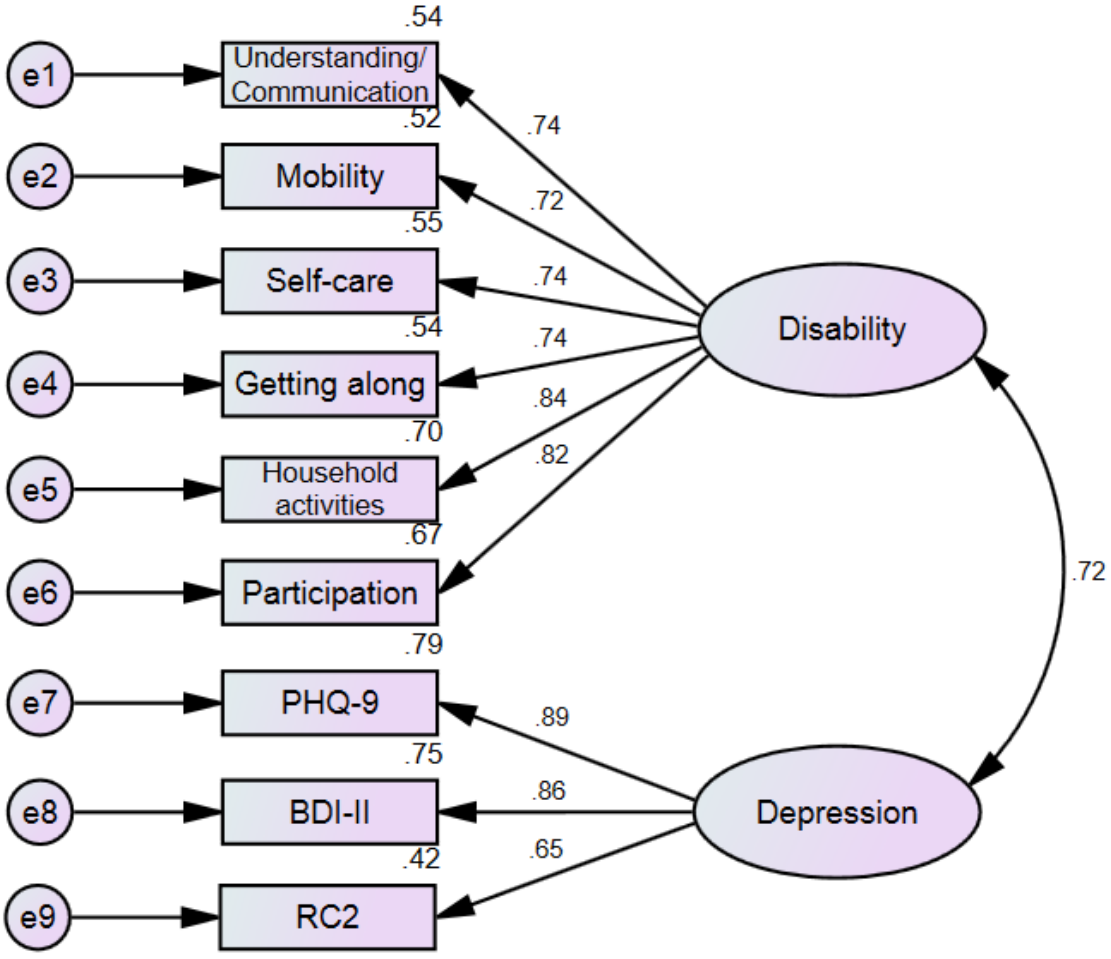

Figure 3. Initial Measurement Model for Latent Variables Disability and Depression. Values to the upper right of the indicators are lower-bound reliability estimates for the

indicator (r^2). Values located immediately above the middle of each directional arrow represent the extent to which each factor loads onto the relevant indicators (β). The correlation between disability and depression is $r = .72$, along the double-headed arrow on the right side of the figure.

Table 3.

Standardized and Unstandardized Coefficients for Measurement Model

Observed variable	Latent construct	β (95% CI)	*B*	*SE*
PHQ-9	Depression	0.89 (.84, .94)	6.24	0.31
BDI-II	Depression	0.86 (.81, .91)	10.39	0.57
RC2	Depression	0.65 (.56, .74)	9.36	0.88
Understanding/ Communication	Disability	0.74 (.66, .81)	15.64	1.07
Mobility	Disability	0.72 (.65, .79)	17.17	1.35
Self-care	Disability	0.74 (.68, .81)	16.65	1.15
Getting along	Disability	0.74 (.66, .81)	19.53	1.32
Household activities	Disability	0.84 (.79, .89)	23.85	1.4
Participation	Disability	0.82 (.77, .87)	18.84	1.07

Note: PHQ-9, Patient Health Questionnaire - 9 item; BDI-II, Beck Depression Inventory, 2nd edition; MMPI-2-RF, Minnesota Multiphasic Personality Inventory, Second edition, Restructured Form; RC2, Restructured Clinical Scale 2 (Low Positive Emotion).

The expanded measurement model is presented in Figure 4. As noted, this model includes the additional measurement components, the latent constructs cognition and symptom validity, and their relationships with one another, as well as disability and depression. Model fit indices essentially fell just within acceptable ranges, $\chi^2(40, N = 282) = 107.70$, $p < .01$, Yuan-Bentler correction factor = 1.12, TLI = 0.92, CFI = 0.94, RMSEA = .08 (90% CI = .06, .10), SRMR = .05.

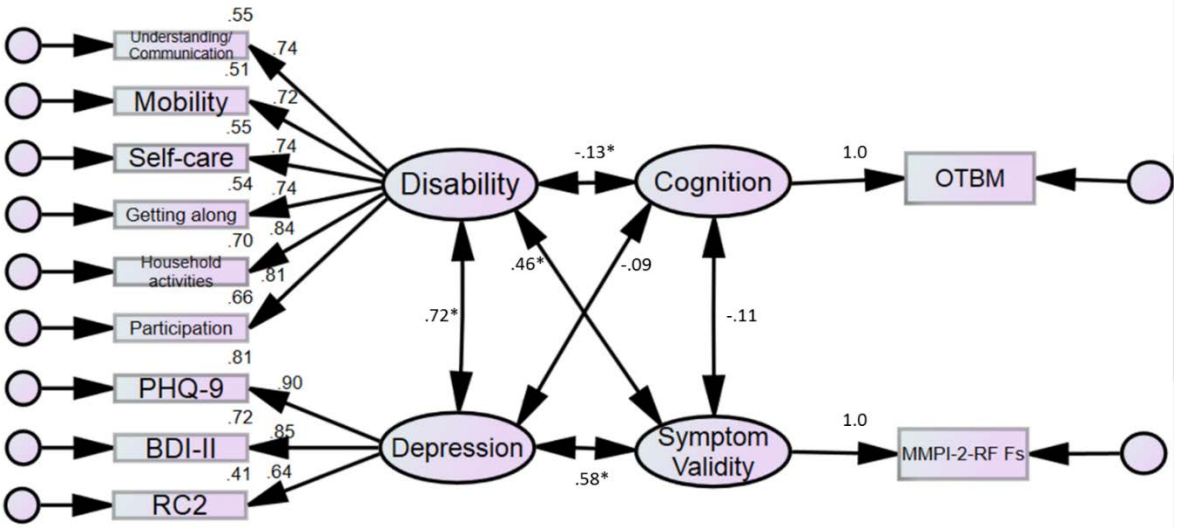

Figure 4. Expanded Measurement Model. Values to the upper right of the indicators are lower-bound reliability estimates for the indicator (r^2). Values located immediately above the head of each directional arrow represent the extent to which each factor loads onto the relevant indicators (β). Double-headed arrows represent correlations between constructs. *$p < .01$.

Table 4 shows standardized parameter estimates with 95% confidence intervals, and unstandardized parameter estimates and standard errors for the model. Significance and magnitude of parameter estimates is approximately identical between the models. The correlation of depression with symptom validity was significant and strong ($r = .46$, $p < .01$), indicating a strong positive association between endorsement of infrequent somatic symptoms and greater level of disability. Notably, the correlation between depression and cognition ($r = -.09$, $p = .15$) was weak and nonsignificant, as was the correlation of cognition with symptom validity ($r = -.11$, $p = .09$).

Table 4.

Standardized and Unstandardized Coefficients for Expanded Measurement Model

Observed variable	Latent construct	β (95% CI)	*B*	*SE*
PHQ-9	Depression	0.90 (.86, .94)	6.31	0.28
BDI-II	Depression	0.85 (.80, .90)	10.23	0.58
RC2	Depression	0.64 (.55, .73)	9.29	0.89
Understanding/ Communication	Disability	0.74 (.66, .81)	15.34	1.10
Mobility	Disability	0.72 (.64, .79)	18.21	1.28
Self-care	Disability	0.74 (.68, .81)	16.84	1.10
Getting along	Disability	0.74 (.66, .81)	19.09	1.33
Household activities	Disability	0.84 (.79, .89)	24.73	1.30
Participation	Disability	0.82 (.76, .87)	18.68	1.06
OTBM*	Cognition	-	7.25	0.32
MMPI-2-RF Fs*	Symptom Validity	-	21.88	0.95

Note: *These parameters were constrained to 1.00 to achieve identification. PHQ-9, Patient Health Questionnaire - 9 item; BDI-II, Beck Depression Inventory, 2nd edition; MMPI-2-RF, Minnesota Multiphasic Personality Inventory, Second edition, Restructured Form; RC2, Restructured Clinical Scale 2 (Low Positive Emotion); OTBM, Overall Test Battery Mean; Fs, MMPI-2-F Infrequent Somatic Complaints scale.

Structural models. The measurement model described above served as the basis for the preliminary structural model in order to establish interrelations among latent constructs and observed variables for the prediction of disability. Model fit was generally acceptable, $\chi^2(40, N = 282) = 107.70$, $p < .01$, Yuan-Bentler correction factor = 1.12, TLI = 0.92, CFI = 0.94, RMSEA = .08 (90% CI = .06, .10), SRMR = .05. Standardized parameter estimates are graphically described in Figure 5. Standardized parameter estimates, their 95% confidence intervals, and unstandardized parameter estimates can be found in Table 5.

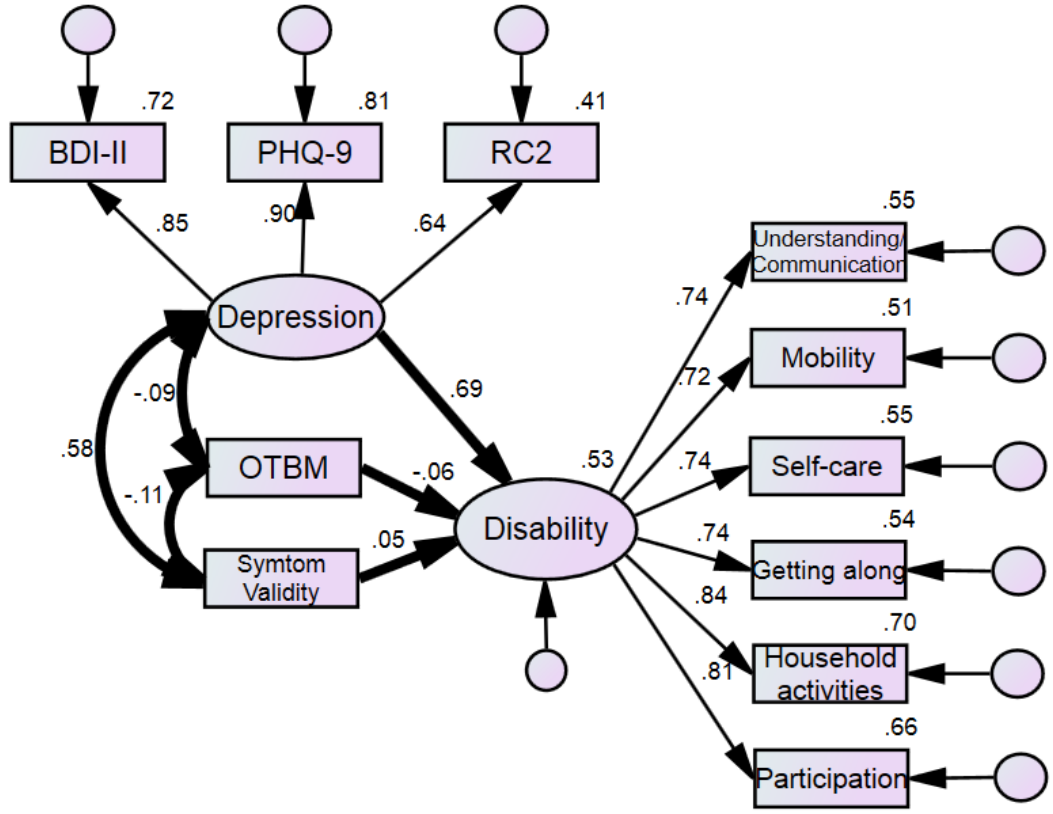

Figure 5. Results of the preliminary structural equation model.

Table 5.

Standardized and Unstandardized Coefficients for Preliminary Structural Model

	β (95% CI)	*B*	*SE*
Depression	.69 (.58, .80)	1.00	0.13
Cognition	-.06 (-.16, .04)	-0.01	0.01
Symptom Validity	.05 (-.08, .18)	0.00	0.00

Note. $R^2 = 0.53$.

Depression was strongly positively related to self-reported disability (β = .69, $p <$ .005). Consistent with prior studies, it was predicted that objective cognitive performance would be weakly related to disability. As anticipated, standardized coefficients were weaker, and in the expected direction (better cognitive performance would be associated

with a lesser degree of disability), although the relationship was not statistically significant (β = -.06, p = .26). Symptom validity was a weak predictor of disability, although the relationship was, as expected, positive (β = .05, p = .46).

When the predictor representing the interaction term was added to the model, in order to test the moderation hypothesis, model fit was consistent, $\chi^2(48, 282) = 113.77$, $p < .01$, Yuan-Bentler correction factor = 1.10, TLI = 0.93, CFI = 0.95, RMSEA = .07 (90% CI = .06, .09), SRMR = .05. Parameter estimates are displayed in Table 6. Notably, the interaction of cognitive performance with symptom validity was a weak predictor of disability, and statistically nonsignificant (β = .03, p = .59). The magnitude and significance of the other predictors was relatively unchanged. Total variance predicted by the model was $R^2 = 0.53$.

Table 6.

Standardized and Unstandardized Coefficients for Interaction Structural Model

	β (95% CI)	*B*	*SE*
Depression	.69 (.57, .80)	1.00	0.13
Cognition	-.06 (-.16, .04)	-0.01	0.01
Symptom Validity	.05 (-.08, .18)	0.00	0.01
OTBM x Symptom Validity	.03 (-.07, .13)	0.04	0.07

Note. $R^2 = 0.53$.

In an attempt to better understand the functioning of symptom validity in its role as a potential predictor and moderator, supplementary exploratory analyses were conducted. The first of these analyses dichotomized Symptom Validity according to clinical cutoffs (MMPI-2-RF Fs *t*-score > 99), and incorporated this new variable into the model. Of the 242 subjects for whom SVT performance was available, 204 (84.3%) produced a “valid” profile, and the remaining 38 (15.7%) scored above the recommended

cutoff, suggesting invalidity of responses. The interaction term also incorporated this dichotomized invalidity measure. The fit of the structural model excluding the interaction term was notably worse, and below acceptable values, as apparent in fit indices, $\chi^2(41, 282) = 150.78$, $p < .01$, Yuan-Bentler correction factor = 1.10, TLI = 0.88, CFI = 0.91, RMSEA = .10 (90% CI = .08, .12), SRMR = .10. Fit was comparable when the interaction term was added to the expanded model, $\chi^2(49, 282) = 158.45$, $p < .01$, Yuan-Bentler correction factor = 1.07, TLI = 0.89, CFI = 0.91, RMSEA = .09 (90% CI = .08, .11), SRMR = .10. Standardized and unstandardized parameter estimates for both models are presented in Table 7. Importantly, the relative statistical significance of parameter estimates in these adjusted models did not change, and the main effects were roughly consistent with previous models. Note that in this model, SVT failure was '1', such that a positive regression weight suggests that those who failed symptom validity endorsed greater disability. The interaction term, in this case, was nonsignificant and small in magnitude.

Table 7.

Standardized and Unstandardized Coefficients for Preliminary and Expanded Structural Models with Dichotomized Symptom Validity Variable

		β (95% CI)	*B*	*SE*
Main Effect Model	Depression	.71 (.62, .80)	1.02	0.13
	Cognition	-.06 (-.17, .04)	-0.01	0.01
	Symptom Validity	.03 (-.09, .14)	0.10	0.24
R^2 = .52				
Interaction Effect Model	Depression	.71 (.62, .80)	1.02	0.13
	Cognition	-.07 (-.19, .05)	-0.01	0.01
	Symptom Validity	.03 (-.09, .15)	0.11	0.24
	OTBM x Symptom Validity	.01 (-.10, .12)	0.04	0.19
R^2 = .51				

Note. When dichotomized, symptom validity failure = 1, such that the relationship between endorsement of infrequent somatic symptoms and disability was expected to be positive.

It was suspected that the strong predictive power of depression could overshadow the relatively smaller potential effect of the hypothesized moderation. To fully address this question, depression was removed from the existing models and they were reanalyzed, in further exploratory analysis. The resulting measurement model showed acceptable fit, $\chi^2(9, N = 260) = 24.97$, $p < .05$, Yuan-Bentler correction factor = 1.29, TLI = 0.957, CFI = 0.974, RMSEA = .094 (90% CI = .051, .139), SRMR = .033. Parameter estimates for the indicators are presented in Table 8.

Table 8.

Standardized and Unstandardized Coefficients for Single-Factor (Disability) Model

Observed variable	β (95% CI)	*B*	*SE*
Understanding/ Communication	.70 (.62, .78)	14.39	1.11
Mobility	.73 (.66, .80)	18.46	1.22
Self-care	.76 (.71, .82)	17.15	1.05
Getting along	.71 (.62, .79)	18.2	1.35
Household activities	.86 (.81, .90)	25.05	1.23
Participation	.80 (.74, .85)	18.04	1.08

Having achieved acceptable fit, no modifications were made prior to fitting the structural models. The main effect structural model revealed acceptable fit, $\chi^2(19, 278) = 48.17$, $p < .01$, Yuan-Bentler correction factor = 1.17, TLI = 0.94, CFI = 0.96, RMSEA = .08 (90% CI = .05, .11), SRMR = .04. Excluding depression as a predictor did not significantly change the magnitude of the regression coefficient for cognition (β = -.09, $p = .16$), but the parameter estimate for symptom validity was notably larger, and was significant (β = .43, $p < .005$). Total variance predicted by the model was $R^2 = .20$, much small than the variance predicted by the model which included depression ($R^2 = .51$). Adding the interaction term revealed a model with approximately equivalently acceptable fit, $\chi^2(25, 278) = 55.21$, $p < .01$, Yuan-Bentler correction factor = 1.13, TLI = 0.94, CFI =

0.96, RMSEA = .07 (90% CI = .05, .10), SRMR = .04. The interaction term was not a statistically significant predictor, and its magnitude was small (β = .07, p = .26). Total variance explained by this model was R^2 = .20. Parameter estimates are presented in Table 9.

Table 9.

Standardized and Unstandardized Coefficients for Main Effect and Interaction Structural Models Excluding Depression

		β (95% CI)	*B*	*SE*
Main Effect Model	Cognition	-.09 (-.20, .03)	-0.01	0.01
	Symptom Validity	.43 (.31, .54)	0.02	0.00
R^2 = .20				
Interaction Effect Model	Cognition	-.09 (-.21, .02)	-0.01	0.01
	Symptom Validity	.43 (.31, .55)	0.02	0.00
	OTBM x Symptom Validity	.07 (-.06, .20)	0.08	0.07
R^2 = .20				

An additional supplemental exploratory analysis involved dichotomizing symptom validity performance and creating two groups of subjects, those that passed versus failed the SVT. This was conducted to rule out the potential differential functioning of the independent relationship between cognition and self-reported disability in terms of SVT performance. Correlations were nonsignificant, with the singular exception of the correlation between OTBM and the understanding/ communication subdomain of WHODAS 2.0. Notably, this subdomain is also occasionally referred to as the “cognitive” domain, as the items are broadly linked to daily cognitive functioning. Results are presented in Table 10.

Table 10.

Correlations Between OTBM and WHODAS Subdomains and Total Score

SVT Group	Understanding/ Communication	Mobility	Self-care	Getting along	Household activities	Participation	Total score
Pass (Fs < 100)	-0.29*	0.06	-0.15	0.13	0.00	0.09	-0.06
Fail (Fs ≥ 100)	-0.20	0.14	0.23	-0.08	0.01	-0.21	-0.11

Note. *p < .01. OTBM, Overall Test Battery Mean; WHODAS, World Health Organization Disability Assessment Schedule; SVT, Symptom Validity Test; Fs, MMPI-2-RF Fs Scale (Infrequent Somatic Complaints).

As previously discussed, samples similar to this are notable and unique for the relatively high rate of performance validity failures. The current sample is consistent with this expectation. Thirteen subjects (4.6%) did not complete the WMT (P. Green, 2005), and of those that did, 109 (40.1%) performed below recommended cutoffs, suggesting performance invalidity. The remaining 163 (59.9%) scored above recommended cutoffs, and had performance that was considered valid. The high rate of PVT failure in the sample suggests another factor by which the expected model relationships may vary. As such, the sample was dichotomized on PVT pass/failure and the model reanalyzed to again examine the differences in the predictive power of depression, cognition, symptom validity, and the interaction of cognition and symptom validity. Model fit of the main effects model, which excluded the interaction term, was borderline acceptable, $\chi^2(84, 272) = 174.94$, $p < .01$, Yuan-Bentler correction factor = 1.06, TLI = 0.91, CFI = 0.93, RMSEA = .09 (90% CI = .07, .11), SRMR = .06. Depression was the only statistically significant predictor of disability in subjects who passed PVT, with a large effect size ($\beta = .65$, $p < .01$). Again, cognition and symptom validity were statistically nonsignificant predictors, and small in magnitude. Symptom validity was a relatively stronger predictor of disability than cognition in the PVT pass group. Notably, the coefficients were in the

expected direction (that is, greater cognitive performance had a negative relationship with disability, while symptom validity had a positive relationship with reported disability). Examining the PVT failure group, depression was the only statistically significant predictor of disability (β = .71, $p < .01$). In this group, symptom validity showed a relatively lower magnitude coefficient, in the opposite direction as predicted (i.e., negative; β = -.08, $p = .49$). Fit of the interaction structural model was also borderline acceptable, $\chi^2(100, 272) = 195.95$, $p < .01$, Yuan-Bentler correction factor = 1.03, TLI = 0.90, CFI = 0.93, RMSEA = .09 (90% CI = .07, .10), SRMR = .06. Adding the interaction effect to the model resulted in almost identical regression weights and did not increase total variance in disability predicted. Parameter estimates are shown in Table 11.

Table 11.

Standardized and Unstandardized Coefficients for Structural Equation Model by PVT Group

		PVT Pass			PVT Fail		
		β (95% CI)	*B*	*SE*	β (95% CI)	*B*	*SE*
Main Effects Model	Depression	.65 (.49, .80)	0.95	0.18	.71 (.53, .90)	0.96	0.20
	Cognition	-.03 (-.15, .08)	-0.01	0.01	.01 (-.19, .20)	0.00	0.14
	Symptom Validity	.13 (-.03, .29)	0.01	0.01	-.08 (-.32, .15)	-0.01	0.03
		R^2 = .54			R^2 = .45		
Interaction Effects Model	Depression	.65 (.49, .80)	0.95	0.18	.71 (.52, .89)	0.96	0.20
	Cognition	-.03 (-.15, .08)	-0.01	0.01	-.01 (-.22, .20)	-0.00	0.02
	Symptom Validity	.13 (-.03, .29)	0.01	0.01	-.07 (-.31, .17)	-0.01	0.01
	OTBM x Symptom Validity	-.02 (-.14, .10)	-0.03	0.09	.03 (-.19, .25)	0.04	0.15
		R^2 = .54			R^2 = .46		

Note. PVT, Performance Validity Test; OTBM, Overall Test Battery Mean

Depression was then removed from the models. The main effects model fit was borderline acceptable, $\chi^2(38, 270) = 76.65$, $p < .01$, Yuan-Bentler correction factor = 1.12, TLI = 0.93, CFI = 0.95, RMSEA = .09 (90% CI = .06, .12), SRMR = .05. Excluding depression, symptom validity was a statistically significant and strong predictor of disability in the PVT passing group (β = .52, $p < .005$). Total variance

predicted was 28%. In subjects who failed PVT, symptom validity was a marginally significant predictor with moderate strength (β = .26, p = .06). Notably, this model predicted only 7% of the variance in disability. The model fit for the interaction model was acceptable, $\chi^2(50, 270) = 92.43$, $p < .01$, Yuan-Bentler correction factor = 1.07, TLI = 0.92, CFI = 0.95, RMSEA = .08 (90% CI = .06, .11), SRMR = .05. Results were similar to those revealed by the main effects model, and are described in Table 12.

Table 12.

Standardized and Unstandardized Coefficients for Structural Models Excluding Depression, by PVT Group

		PVT Pass			PVT Fail		
		β (95% CI)	*B*	*SE*	β (95% CI)	*B*	*SE*
Main Effects Model	Cognition	-.02 (-.16, .13)	-0.00	0.01	.03 (-.18, .23)	0.00	0.02
	Symptom Validity	.52 (.41, .64)	0.03	0.00	.26 (.01, .51)	0.01	0.01
		R^2 = .28			R^2 = .07		
Interaction Effects Model	Cognition	-.02 (-.16, .13)	-0.00	0.01	.01 (-.20, .21)	0.00	0.02
	Symptom Validity	.52 (.40, .65)	0.03	0.00	.28 (.00, .55)	0.01	0.01
	OTBM x Symptom Validity	.01 (-.15, .17)	0.01	0.10	.05 (-.18, .28)	0.05	0.11
		R^2 = .28			R^2 = .07		

Note. PVT, Performance Validity Test; OTBM, Overall Test Battery Mean

General findings from these supplementary analyses demonstrated that in the context of PVT performance that is within the normal range, symptom validity is more strongly related to self-report of disability than in subjects whose PVT performance is considered sub-optimal. In both patient groups, cognitive test performance itself was only minimally predictive of self-reported disability. The addition of the interaction term did not significantly contribute to the prediction of disability, in any variation of the analyses.

Aim 2

Despite nonsignificant findings in the moderation hypothesis as previously described, further investigation was undertaken to determine if, when dichotomized, there were

group differences evident in levels of endorsement on self-report measures between those who passed versus failed SVT. It was hypothesized that WHODAS 2.0 Total score would be significantly higher in those who failed SVT as compared to those who did not. By extension, it stands to reason that WHODAS subdomains may also vary in this way, though specific predictions were not made. Due to non-normality in most of the self-report variables, and violation of other assumptions in analysis of variance, such as homogeneity of variances, nonparametric tests for group differences were used with the exception of WHODAS 2.0 Total score, which was normally distributed. Specifically, the Mann-Whitney *U* Test (Mann & Whitney, 1947), which is a rank-based test for a dichotomous independent variable and continuous or ordinal dependent variable. Distributions of scores on each dependent variable were visually inspected and all were found to be dissimilar. Results are presented in Table 13. As predicted, WHODAS 2.0 Total score was significantly different between those who passed (M = 36.40, SD = 18.16) and failed (M = 50.44, SD = 18.42) SVT, $F(1, 234) = 18.13$, $p < .005$, with a moderate effect size (d = .77). While scores for the SVT failure group were significantly higher on all self-report measures compared ($p < .05$), effect sizes ranged from small (Mobility, Self-care, Household activities, Participation) to moderate (MMPI-2-RF RC2, Understanding/ Communication, BDI-II, Getting along, PHQ-9). A strong effect size was noted only for Fs (r = .63). These findings are consistent with previously published findings that endorsement on self-report measures is, to some extent, predicted by whether the examinee is responding in an accurate/truthful way.

Table 13.

Mann-Whitney U Tests of Symptom Validity Group Median Differences

		SVT Pass		SVT Fail					
Variable	*N*	Median	Mean rank	Median	Mean rank	*U*	*z*	*p*	*r*
MMPI-2-RF Fs	242	74.39	102.50	115.00	223.50	7752.00	9.80	< .005	.63
PHQ-9	236	10.00	106.15	18.00	187.11	6070.00	6.56	< .005	.43
BDI-II	241	18.00	112.34	25.50	167.26	5615.00	4.46	< .005	.29
MMPI-2-RF RC2	242	61.00	114.42	74.50	159.53	5321.00	3.66	< .005	.24
Understanding/ Communication	239	30.00	112.13	55.00	162.97	5327.00	4.12	< .005	.27
Mobility	239	43.75	115.62	50.00	143.93	4622.50	2.30	.022	.15
Self-care	239	10.00	115.33	30.00	145.47	4679.50	2.49	.013	.16
Getting along	238	33.33	111.02	58.33	165.57	5423.00	4.45	< .005	.29
Household activities	238	40.00	114.32	50.00	148.58	4683.00	2.78	.005	.18
Participation	236	45.83	112.25	60.42	153.24	4850.50	3.32	.001	.22

Note: SVT, Symptom Validity Test; Fs, MMPI-2-RF Infrequent Somatic Complaints scale; PHQ-9, Patient Health Questionnaire - 9 item; BDI-II, Beck Depression Inventory - 2nd edition; RC2, MMPI-2-RF Restructured Clinical Scale 2 (Low Positive Emotion); WHODAS, World Health Organization Disability Assessment Schedule

Having established the above differences in item endorsement between SVT groups, various demographics were investigated for their association with SVT performance. Analysis of variance (ANOVA) revealed no difference in age between subjects failing (M = 47.00, SD = 13.34) versus passing (M = 50.03, SD = 14.43) SVT, $F(1, 240) = 1.44$, $p = .23$. Likewise, there was no difference in educational attainment between SVT pass and fail groups (M = 13.28, SD = 2.25 and M = 12.71, SD = 2.62, respectively), $F(1,235) = 1.91$, $p = .17$. Regarding sex, of the 58 females and 184 males, 19.0% and 14.7% failed SVT, respectively. Chi square analysis revealed no association between SVT group and sex, $\chi^2(1, 242) = .61$, $p = .43$. In order to investigate the association of SVT group with ethnicity, only the two highest populated ethnic groups (African American and Caucasian) were used, due to very low cell counts in the other three groups. Chi square analysis using all ethnic groups would yield a poor approximation of the chi square distribution. Comparing African American and Caucasian groups, 24 (14.5%) of the 165 Caucasians, and 13 (21.7%) of the 60 African-Americans failed SVT. There was not a significant association between SVT group and ethnicity that emerged, considering these ethnic groups, $\chi^2(1, 225) = 1.62$, $p = .20$. Considering diagnosis, several diagnostic groups had very low cell counts. As before, the largest two groups (PNES and epilepsy) were considered for chi square analysis, to avoid poor approximation of the distribution. The modified chi square revealed a significant association between diagnosis and SVT group, $\chi^2(1, 150) = 5.21$, $p < .05$. Specifically, a greater proportion of subjects with PNES failed SVT, compared to subjects with epilepsy (21.9% versus 7.4%).

Discussion

Aim 1

It was hypothesized that in the interrelations between cognition, symptom validity, and self-reported disability, that symptom validity would moderate the relationship between cognition and self-reported disability. Such a predictive relationship has not been demonstrated in the literature previously, however it does provide a unique and theory-based method of examining these relationships. As previously discussed, cognition has not been strongly related to self-report of functional limitations and disablement. Rather, objective cognitive performance is generally more strongly associated with measures of capacity (that is, functional objective measures that require the completion of a task). Self-reported disability and limitations are instead more strongly associated with the presence and severity of mood and psychiatric symptoms, such as depression.

The current study addresses the issue of symptom validity, which has been shown to be strongly associated with levels of symptoms reported. Given the need to understand how WHODAS 2.0 is influenced by symptom validity (e.g., as described in Gold, 2014), the theoretical model of interest in the present study addressed the intersection of objective cognitive performance, symptom validity, depression, and self-report of disability.

The strong psychometric properties of the measures used helped in establishing an adequate measurement model for confirmatory factor analysis. Data which were found to be non-normally distributed necessitated the use of statistical approaches, such as MLR, which result in robust and unbiased parameter estimates. As such, the findings from the

present study are thought to be reasonably representative of the population interrelations of the variables of interest.

A measurement model with adequate performance according to common fit indices was confirmed prior to testing the more complicated structural model. Factor loadings of the CFA were moderately strong, and there was a strong relationship between the constructs of interest, depression and disability. The follow-up measurement model, which related all constructs of interest to one another by adding cognition and symptom validity, revealed additional strong associations of symptom validity with depression and disability. As previously stated, this is consistent with general findings in other populations using similar measures.

Predictive pathways were incorporated to test the structural model and the proposed moderation hypothesis. Model fit was moderately acceptable overall. Findings from the SEM did not support the hypothesized moderating effect of symptom validity on the relation between cognition and self-reported disability. Unsurprisingly, depression was by far the largest, and only statistically significant predictor of disability, accounting for the majority of variance explained in the models. Reasons for the findings are likely multifactorial, including limitations discussed below, as well the nature of the symptom validity scale chosen. The majority of symptom validity literature has conflated the constructs of symptom over-reporting/exaggeration, malingering, and performance validity. As such, the Fs scale, which has good specificity in detecting “malingering” in forensic samples, may be less sensitive in a mixed clinical non-litigating population, such as the EMU sample in the present study. This may also account for the lower than

expected correlations between Fs and WHODAS 2.0 subdomain scores. It may also be the case that meaningful factors or measurement and structural elements that were not considered in the study may influence the relationships of interest in the current model. These factors might include, for example, perceived seizure severity, medical and psychiatric comorbidities, or the presence of physical disabilities. When depression was removed from the model, symptom validity emerged as a significant predictor of disability, although it was a weaker predictor. Interestingly, in the model excluding depression, while comparing groups of subjects who passed versus failed PVT, there was a significant difference in variance in disability explained between the two groups. Symptom validity was a much stronger predictor of disability in those who passed PVTs. This seems to suggest that (a) there may be more variability in responding evident in subjects failing PVT (inconsistency in responding), and (b) performance validity may be the factor more likely to have a noticeable effect in similar models. Such an interpretation would be consistent with previous literature, and merit further study of the phenomenon in other samples.

Aim 2

The secondary aim of the study was to examine differences in self-reported disability as a function of SVT response style (valid versus invalid). Consistent with prior studies, WHODAS 2.0 scores were higher in participants who failed SVT. As a supplement to the SEM analysis conducted to address the study's primary aims, group differences provide an easily interpretable and clinically useful method for understanding the functioning of subjective measures, including perceived functional limitation. Scores on other self-

report measures, including all depression scales examined, were significantly higher in those who failed SVT. Such findings suggest that patients who exceed the recommended cutoff on Fs are likely to report more symptoms, or greater symptom severity, than their counterparts who pass SVT. Whether the increased symptom endorsement is the result of over-reporting or exaggeration *per se*, as opposed to the genuine presence of greater pathology or impairment is a matter of interpretation. An unequivocal determination could only be made by verifying the presence of the symptoms themselves. If, however, we are to assume that Fs is, in fact, sensitive and specific to deliberate exaggeration, we may be more confident inferring that the concurrent elevations indeed suggest exaggeration on those measures as well.

Contributions of the present study

This study was conceived to investigate a relatively recently developed and adopted self-report measure of functional limitations, the WHODAS 2.0, with a particular interest in its relation to objective cognitive functioning, as well as symptom validity. WHODAS 2.0 has been widely studied and validated in terms of its ability to measure perceived disability, although the existing literature has broadly failed to take into account the validity of symptom reporting, and has more commonly focused on the relationship between disability, cognition, and mood. Given the prevalence of symptom exaggeration in many clinical settings, the utility of WHODAS 2.0 would be improved by understanding how it is influenced by the accuracy of symptom reporting.

The use of structural equation modeling in the research design served to improve the robustness of the study and the interpretation of the findings herein. SEM is a

powerful statistical tool that has been applied to a wide range of research questions, though has been largely absent from published literature in neuropsychological and disability assessment. An extensive body of research developing and validating robust procedures like those used in this study (e.g., robust maximum likelihood estimation) give researchers tools to adequately address common issues in social science research (e.g., missing data, non-normal data). Requiring adherence to relatively fewer statistical assumptions is likely to expand the range of questions that can be asked, while improving the quality of their answers.

Despite the failure to detect a significant effect of symptom validity on the relationship in question, this study still contributes important information to the understanding of these relationships. By replicating the finding that depression is the factor that is most predictive of self-reported disability, the present study has extended this known relationship to a mixed sample of patients with epileptic seizures, psychogenic nonepileptic events, and other medical and psychiatric conditions. Perceived disability is, itself, a meaningful clinical variable worth consideration in the context of neuropsychological evaluation.

Limitations

As previously mentioned, the use of SEM and robust parameter estimation procedures is a clear strength of the present study, particularly in a literature that rarely if ever applies SEM to these relationships. However, the study is not without its limitations. Perhaps the most significant of which is sample size. While many basic statistical procedures require relatively few participants to achieve adequate statistical power, complex models such as

those investigated in this study require substantially greater numbers of participants. Considering issues specific to this study, exploratory SEM analyses were especially likely to be underpowered to detect especially small effects, due to increasing model complexity, and subdividing the sample into two groups. General "rules of thumb" for minimum sample size in CFA and SEM are non-specific and tend to ignore important model characteristics such as the number of latent constructs in the model, the factor loadings of indicators, the proportion of missing data, and the number of parameters being estimated (Wolf et al., 2013). As previously mentioned, a power analysis conducted for one of the model fit indices of interest, RMSEA, revealed that the study may have insufficient power, and require a calculated 367 participants (82 additional subjects; a sample size increase of 28.8%) to achieve adequate power. Based on simulations conducted by Wolf and colleagues, and given this study's two-factor CFA, using the lower bound of three indicators per factor, assuming conservatively estimated moderate factor loadings (i.e., .65, approximately consistent with the experimental data), the minimum sample size is likely to fall near $N = 200$. The discrepancy in this estimation versus the power analysis calculation for RMSEA could be the result of failure to consider the number of parameters being estimated, or the correlations among indicators in the model, for instance.

Due to the retrospective nature of this study, data available for consideration in the model was limited to that collected for clinical purposes in the EMU. A number of transdiagnostic characteristics that could be relevant to the current study were unavailable. Although most basic demographic variables were included and described for

the present sample, these did not include such potentially relevant variables such as medical and psychiatric comorbidities present, number and type of psychoactive medications including anti-epileptic drugs (AEDs), estimates of premorbid IQ, service-connection status, compensation-seeking status, or employment status. Further, the neuropsychological testing administered in the EMU is limited to brief assessments, as patients are frequently unavailable for extended periods of time as they undergo other time-consuming evaluations such as MRI. While the cognitive test variables available provide a reasonable range of cognitive domains sampled, each domain is not thoroughly evaluated. This raises two separate concerns. First, are the cognitive domains most closely linked to functional limitations being adequately assessed? For example, executive functioning has been linked to disability (Kiosses, Klimstra, Murphy, & Alexopoulos, 2001), and verbal memory deficits common in temporal lobe epilepsy (Äikiä et al., 2001) are likely to impact daily functioning as well. However, only one "true" verbal memory measure is administered to patients in the EMU (WMS-IV Logical Memory). The other procedures assessing memory are the performance validity tests which use a recognition memory paradigm, with psychometric properties that are not conducive to inclusion in a cognitive test composite as was used in this study. The brief neurocognitive battery that patients are administered contains no formal measures of language or visuospatial abilities. The second concern, pervasive in the literature on neuropsychological testing, is how do we interpret "impure" measures of the domains we are assessing? Part B of the Trail Making Test was the only measure of executive functioning (EF) in the present study, however it cannot be considered a "pure" test of

EF. The test is timed, therefore confounded by processing speed, in addition to being dependent on motor functioning and visual abilities, none of which are exclusively "executive functions." These examples are representative of broad issues with retrospective methodologies, and could be extended to critique the selection of any variables in this study. Similarly, knowing that subjective measures of disability are less likely to be related to objective measures of cognition, future studies may relate objective cognitive testing to measures of functional *capacity.* That is, the patient's *actual* rather than *perceived* ability.

The sample in question was composed entirely of Veterans, mostly male, with very few subjects who were not Caucasian or African-American. Thus, the results of the current study may be limited in generalizability. The sample was further restricted to those undergoing evaluation for epilepsy and seizure disorders, itself a unique population due to high rates of trauma, medical and psychiatric comorbidities, and of course the presence of functional neurological symptom disorder and possible symptom exaggeration. Each of these on its own provides an opportunity to answer an important question in a unique population. Unfortunately, due to statistical power considerations, it was not possible to further subdivide the sample, and as such this is considered a limitation.

Generalizability is also limited in the current study because a diagnosis of epilepsy is an exclusionary criteria for serving in the military. As such, all of the patients in the current sample with epilepsy have a relatively later age of onset (after reaching adulthood). Because of the effects of ictal and interictal epileptiform discharges present

during sensitive neurodevelopmental periods, patients with earlier onset of epilepsy are likely to exhibit greater cognitive impairment relative to those with later onset (Van Rijckevorsel, 2006). Subjects in the current sample with epilepsy may have performed better than their counterparts with earlier onset of epilepsy, and may have different levels of perceived disability.

Future Directions

Considering the findings of this study, in the context of similar work that has been done across other populations, the limitations mentioned above provide areas of development for future research. For instance, simply having a much larger sample would improve statistical power and raise confidence in the results obtained. Further, the ability to subdivide the sample and compare model statistics between groups (i.e., diagnostic, demographic, or other groups) would provide valuable clinical information. For example, a similar study may compare model fit and parameter estimates between subjects with epileptic seizures and those with psychogenic nonepileptic events.

The elevation noted on WHODAS 2.0 in the SVT failure group suggests that, as Gold (2014) described, the measure is subject to the same mechanisms of increased reporting as symptom validity and symptom severity scales. This finding further supports the need for establishing empirically-derived cutoff scores, above which item endorsement on WHODAS 2.0 may reflect over-reporting of functional limitations. Gold additionally expressed criticism that WHODAS 2.0 had been recommended for psychiatric and medical disability evaluations, given that it has no embedded validity

measure. Future work may include known-groups analyses to determine odds ratios for clinical cutoff scores that neuropsychologists or others may find useful.

Using a large enough sample, conducting an exploratory factor analysis (EFA) to investigate the relative influence of other variables and/or model relationships would help to more thoroughly describe the relationships among these variables in various populations. EFA procedures, if conducted in a large enough sample, can be enhanced by performing EFA on half of the sample, then the model can be verified by CFA on the other half of the sample. For instance, some previous models have shown that the relationship between objective cognitive performance and subjective disability is mediated by depression (Ventura, Hellemann, Thames, Koellner, & Nuechterlein, 2009; Weber, Mapstone, Staskiewicz, & Maki, 2012). A potentially interesting variation on this finding could explore a moderated mediation model in which the extent to which depression mediates the relationship is moderated by social support.

References

Äikiä, M., Salmenperä, T., Partanen, K., & Kälviäinen, R. (2001). Verbal memory in newly diagnosed patients and patients with chronic left temporal lobe epilepsy. *Epilepsy & Behavior*, *2*(1), 20–27.

Al Marzooqi, S. M., Baker, G. A., Reilly, J., & Salmon, P. (2004). The perceived health status of people with psychologically derived non-epileptic attack disorder and epilepsy: a comparative study. *Seizure*, *13*(2), 71–75. https://doi.org/10.1016/S1059-1311(03)00158-4

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision). Washington, DC.

American Psychiatric Association. (2013). *The Diagnostic and Statistical Manual of Mental Disorders: DSM 5*. bookpointUS.

Asmussen, S. B., Kirlin, K. A., Gale, S. D., & Chung, S. S. (2009). Differences in self-reported depressive symptoms between patients with epileptic and psychogenic nonepileptic seizures. *Seizure*, *18*(8), 564–566. https://doi.org/10.1016/j.seizure.2009.05.006

Baron, M., Schieir, O., Hudson, M., Steele, R., Kolahi, S., Berkson, L., … Zummer, M. (2008). The clinimetric properties of the world health organization disability assessment schedule II in early inflammatory arthritis. *Arthritis Care & Research*, *59*(3), 382–390. https://doi.org/10.1002/art.23314

Beck, A., Steer, R., & Brown, G. (1996). Manual: Beck Depression Inventory-II. *TX: The Psychological Corporation*.

Benge, J. F., Wisdom, N. M., Collins, R. L., Franks, R., LeMaire, A., & Chen, D. K. (2012). Diagnostic utility of the Structured Inventory of Malingered Symptomatology for identifying psychogenic non-epileptic events. *Epilepsy & Behavior*, *24*(4), 439–444.

Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: MMPI-2-RF*. NCS Pearson, Incorporated.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Binder, L. M., Kindermann, S. S., Heaton, R. K., & Salinsky, M. C. (1998). Neuropsychologic Impairment in Patients with Nonepileptic Seizures. *Archives of Clinical Neuropsychology*, *13*(6), 513–522. https://doi.org/10.1016/S0887-6177(97)00042-5

Binder, L. M., & Salinsky, M. C. (2007). Psychogenic nonepileptic seizures. *Neuropsychology Review*, *17*(4), 405–412.

Binder, L. M., Salinsky, M. C., & Smith, S. P. (1994). Psychological correlates of psychogenic seizures. *Journal of Clinical and Experimental Neuropsychology*, *16*(4), 524–530.

Bodde, N. M. G., Brooks, J. L., Baker, G. A., Boon, P. A. J. M., Hendriksen, J. G. M., & Aldenkamp, A. P. (2009). Psychogenic non-epileptic seizures—Diagnostic issues: A critical review. *Clinical Neurology and Neurosurgery*, *111*(1), 1–9. https://doi.org/10.1016/j.clineuro.2008.09.028

Bombin, I., Santiago-Ramajo, S., Garolera, M., Vega-González, E. M., Cerulla, N., Caracuel, A., … Bobes, J. (2012). Functional impairment as a defining feature of: amnestic MCI cognitive, emotional, and demographic correlates. *International Psychogeriatrics*, *24*(09), 1494–1504.

Bowie, C. R., Depp, C., McGrath, J. A., Wolyniec, P., Mausbach, B. T., Thornquist, M. H., … Pulver, A. E. (2010). Prediction of real-world functional disability in chronic mental disorders: a comparison of schizophrenia and bipolar disorder. *American Journal of Psychiatry*.

Breier, J., Fuchs, K., & Brookshire, B. (1998). Quality of life perception in patients with intractable epilepsy or pseudoseizures. *Archives of Neurology*, *55*(5), 660–665. https://doi.org/10.1001/archneur.55.5.660

Browne, M. W., Cudeck, R., Bollen, K. A., & Long, J. S. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136–136.

Brown, M., Levin, B. E., Ramsay, R. E., Katz, D. A., & Duchowny, M. S. (1991). Characteristics of patients with nonepileptic seizures. *Journal of Epilepsy*, *4*(4), 225–229.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology*, *107*(2), 179.

Buist-Bouwman, M. A., Ormel, J., Graaf, R. D., Vilagut, G., Alonso, J., Sonderen, E. V., & Vollebergh, W. A. M. (2008). Psychometric properties of the World Health

Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders. *International Journal of Methods in Psychiatric Research*, *17*(4), 185–197. https://doi.org/10.1002/mpr.261

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*(06), 547–558.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., … Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426.

Butcher, J. N., Arbisi, P. A., Atlis, M. M., & McNulty, J. L. (2008). The construct validity of the Lees-Haley Fake Bad Scale: Does this scale measure somatic malingering and feigned emotional distress? *Archives of Clinical Neuropsychology*, *23*(7–8), 855–864. https://doi.org/10.1016/j.acn.2008.10.001

Cankurtaran, E. S., Ulug, B., Saygi, S., Tiryaki, A., & Akalan, N. (2005). Psychiatric morbidity, quality of life, and disability in mesial temporal lobe epilepsy patients before and after anterior temporal lobectomy. *Epilepsy & Behavior*, *7*(1), 116–122. https://doi.org/10.1016/j.yebeh.2005.03.019

Carlozzi, N., Kratz, A., Downing, N., Goodnight, S., Miner, J., Migliore, N., & Paulsen, J. (2015). Validity of the 12-item World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) in individuals with Huntington disease

(HD). *Quality of Life Research*, *24*(8), 1963–1971. https://doi.org/10.1007/s11136-015-0930-x

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217–227. https://doi.org/10.1016/j.acn.2005.12.002

Chisolm, T. H., Abrams, H. B., McArdle, R., Wilson, R. H., & Doyle, P. J. (2005). The WHO-DAS II: psychometric properties in the measurement of functional health status in adults with acquired hearing loss. *Trends In Amplification*, *9*(3), 111–126.

Chopra, P. K., Couper, J. W., & Herrman, H. (2004). The assessment of patients with long-term psychotic disorders: application of the WHO Disability Assessment Schedule II. *The Australian And New Zealand Journal Of Psychiatry*, *38*(9), 753–759.

Chwastiak, L. A., & Von Korff, M. (2003). Disability in depression and back pain: Evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. *Journal of Clinical Epidemiology*, *56*(6), 507–514. https://doi.org/10.1016/S0895-4356(03)00051-9

Clark, A. L., Amick, M. M., Fortier, C., Milberg, W. P., & McGlinchey, R. E. (2014). Poor Performance Validity Predicts Clinical Characteristics and Cognitive Test Performance of OEF/OIF/OND Veterans in a Research Setting. *The Clinical Neuropsychologist*, *28*(5), 802–825.

Clarke, B. M., Upton, A., & Castellanos, C. (2006). Work beliefs and work status in epilepsy. *Epilepsy & Behavior*, *9*(1), 119–125.

Cramer, J. A., Perrine, K., Devinsky, O., Bryant-Comstock, L., Meador, K., & Hermann, B. (1998). Development and cross-cultural translations of a 31-item quality of life in epilepsy inventory. *Epilepsia*, *39*(1), 81–88.

DeCarlo, L. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292.

Dodrill, C. B. (2008). Do patients with psychogenic nonepileptic seizures produce trustworthy findings on neuropsychological tests? *Epilepsia*, *49*(4), 691–695. https://doi.org/10.1111/j.1528-1167.2007.01457.x

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, *2*, 222. https://doi.org/10.1186/2193-1801-2-222

Doyle, M., Flanagan, S., Browne, S., Clarke, M., Lydon, D., Larkin, C., & O'Callaghan, E. (1999). Subjective and external assessments of quality of life in schizophrenia: relationship to insight. *Acta Psychiatrica Scandinavica*, *99*(6), 466–472. https://doi.org/10.1111/j.1600-0447.1999.tb00994.x

Drane, D. L., Williamson, D. J., Stroup, E. S., Holmes, M. D., Jung, M., Koerner, E., … Miller, J. W. (2006). Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia*, *47*(11), 1879–1886.

Dworetzky, B. A., Strahonja-Packard, A., Shanahan, C. W., Paz, J., Schauble, B., & Bromfield, E. B. (2005). Characteristics of Male Veterans with Psychogenic

Nonepileptic Seizures. *Epilepsia*, *46*(9), 1418–1422. https://doi.org/10.1111/j.1528-1167.2005.13004.x

Eichstaedt, K. E., Clifton, W. E., Vale, F. L., Benbadis, S. R., Bozorg, A. M., Rodgers-Neame, N. T., & Schoenberg, M. R. (2014). Sensitivity of Green's Word Memory Test Genuine Memory Impairment Profile to Temporal Pathology: A Study in Patients With Temporal Lobe Epilepsy. *The Clinical Neuropsychologist*, *28*(6), 941–953. https://doi.org/10.1080/13854046.2014.942374

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457.

Ertuğrul, A., & Uluğ, B. (2002). The influence of neurocognitive deficits and symptoms on disability in schizophrenia. *Acta Psychiatrica Scandinavica*, *105*(3), 196–201. https://doi.org/10.1034/j.1600-0447.2002.1020.x

Federici, S., & Meloni, F. (2010). WHODAS II: Disability self-evaluation in the ICF conceptual frame. *International Encyclopedia of Rehabilitation*, 1–22.

Federici, S., Meloni, F., Mancini, A., Lauriola, M., & Olivetti Belardinelli, M. (2009). World Health Organisation Disability Assessment Schedule II: Contribution to the Italian validation. *Disability and Rehabilitation*, *31*(7), 553–564. https://doi.org/10.1080/09638280802240498

Federici, S., Meloni, F., & Presti, A. L. (2009). International literature review on WHODAS II (World Health Organization Disability Assessment Schedule II).

Fox, R. J. (2010). Confirmatory Factor Analysis. In *Wiley International Encyclopedia of Marketing*. John Wiley & Sons, Ltd. Retrieved from http://dx.doi.org/10.1002/9781444316568.wiem02060

Francis, P., & Baker, G. A. (1999). Non-epileptic attack disorder (NEAD): a comprehensive review. *Seizure*, *8*(1), 53–61. https://doi.org/10.1053/seiz.1998.0246

Gallagher, P., & Mulvany, F. (2004). Levels of ability and functioning: using the WHODAS II in an Irish context. *Disability and Rehabilitation*, *26*(9), 506–517.

Garin, O., Ayuso-Mateos, J., Almansa, J., Nieto, M., Chatterji, S., Vilagut, G., … the MHADIE consortium. (2010). Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. *Health and Quality of Life Outcomes*, *8*(1), 51. https://doi.org/10.1186/1477-7525-8-51

Gilliam, F., Hecimovic, H., & Sheline, Y. (2003). Psychiatric comorbidity, health, and function in epilepsy. *Epilepsy & Behavior*, *4*, 26–30.

Gold, L. H. (2014). DSM-5 and the Assessment of Functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online*, *42*(2), 173–181.

Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W., & Wright, B. D. (1993). Performance profiles of the functional independence measure. *American Journal of Physical Medicine & Rehabilitation*, *72*(2), 84–89.

Green, M. F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *The American Journal of Psychiatry*, *153*(3), 321–330.

Green, P. (2005). Word Memory Test for Windows: User's manual and program. *Edmonton: Green's Publishing*.

Grudzinski, A. N., Hakim, Z., Coons, S. J., & Labiner, D. M. (1998). Use of the QOLIE-31 in Routine Clinical Practice. *Journal of Epilepsy*, *11*(1), 34–47. https://doi.org/10.1016/S0896-6974(97)00110-2

Guilera, G., Gómez-Benito, J., Pino, Ó., Rojo, E., Vieta, E., Cuesta, M. J., … Rejas, J. (2015). Disability in bipolar I disorder: The 36-item World Health Organization Disability Assessment Schedule 2.0. *Journal of Affective Disorders*, *174*, 353–360. https://doi.org/10.1016/j.jad.2014.12.028

Hackett, M., Glozier, N., Martiniuk, A., Jan, S., & Anderson, C. (2011). Sydney epilepsy incidence study to measure illness consequences: the SESIMIC observational epilepsy study protocol. *BMC Neurology*, *11*(1), 1–11. https://doi.org/10.1186/1471-2377-11-3

Harden, C. L., Maroof, D. A., Nikolov, B., Fowler, K., Sperling, M., Liporace, J., … Herzog, A. (2007). The effect of seizure severity on quality of life in epilepsy. *Epilepsy & Behavior*, *11*(2), 208–211. https://doi.org/10.1016/j.yebeh.2007.05.002

Hartley, H., & Hocking, R. (1971). The analysis of incomplete data. *Biometrics*, 783–823.

Harvey, P. D. (2010). Cognitive functioning and disability in schizophrenia. *Current Directions in Psychological Science, 19*(4), 249–254.

Harwood, R., & Ebrahim, S. (1995). *Manual of the London handicap scale*. Department of Health Care of the Elderly, University of Nottingham Nottingham.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: demographic corrections, research findings, and clinical applications; with a supplement for the Wechsler Adult Intelligence Scale-Revised (WAIS-R)*. Psychological Assessment Resources.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants 1. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093–1129.

Hermann, B., & Seidenberg, M. (2007). Epilepsy and Cognition. *Epilepsy Currents*, *7*(1), 1–6.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions (Vol. 1, pp. 221–233). Presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

Hudson, M., Steele, R., Taillefer, S., Baron, M., & Canadian Scleroderma Research Group. (2008). Quality of life in systemic sclerosis: Psychometric properties of the World Health Organization Disability Assessment Schedule II. *Arthritis Care & Research, 59*(2), 270–278. https://doi.org/10.1002/art.23343

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

Janca, A., Kastrup, M., Katschnig, H., Lopez-Ibor Jr, J., Mezzich, J., & Sartorius, N. (1996). The World Health Organization Short Disability Assessment Schedule (WHO DAS-S): a tool for the assessment of difficulties in selected areas of functioning of patients with mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, *31*(6), 349–354.

Jokeit, H., & Ebner, A. (2002). Effects of chronic epilepsy on intellectual functions. *Progress in Brain Research*, *135*, 455–463.

Kaye, J. L., Dunlop, B. W., Iosifescu, D. V., Mathew, S. J., Kelley, M. E., & Harvey, P. D. (2014). Cognition, functional capacity, and self-reported disability in women with posttraumatic stress disorder: Examining the convergence of performance-based measures and self-reports. *Journal of Psychiatric Research*, *57*, 51–57. https://doi.org/10.1016/j.jpsychires.2014.06.002

Kenny, D. A. (2012). Structural equation modeling.

Kessler, R. C., Lane, M. C., Shahly, V., & Stang, P. E. (2012). Accounting for comorbidity in assessing the burden of epilepsy among US adults: Results from the National Comorbidity Survey Replication (NCS-R). *Molecular Psychiatry*, *17*(7), 748–759. https://doi.org/10.1038/mp.2011.56

Kim, J.-M., Stewart, R., Glozier, N., Prince, M., Kim, S.-W., Yang, S.-J., … Yoon, J.-S. (2005). Physical health, depression and cognitive function as correlates of

disability in an older Korean population. *International Journal of Geriatric Psychiatry*, *20*(2), 160–167. https://doi.org/10.1002/gps.1266

Kiosses, D. N., Klimstra, S., Murphy, C., & Alexopoulos, G. S. (2001). Executive dysfunction and disability in elderly patients with major depression. *American Journal of Geriatric Psychiatry*, *9*. https://doi.org/10.1097/00019442-200108000-00011

Konecky, B., Meyer, E. C., Marx, B. P., Kimbrel, N. A., & Morissette, S. B. (2014). Using the WHODAS 2.0 to Assess Functional Disability Associated With DSM-5 Mental Disorders. *American Journal of Psychiatry*, *171*(8), 818–820. https://doi.org/10.1176/appi.ajp.2014.14050587

Krawetz, P., Fleisher, W., Pillay, N., Staley, D., Arnett, J., & Maher, J. (2001). Family Functioning in Subjects with Pseudoseizures and Epilepsy. *The Journal of Nervous and Mental Disease*, *189*(1). Retrieved from http://journals.lww.com/jonmd/Fulltext/2001/01000/Family_Functioning_in_Subjects_with_Pseudoseizures.7.aspx

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. *Journal of General Internal Medicine*, *16*(9), 606–613.

Küçükdeveci, A. A., Kutlay, S., Yildizlar, D., Öztuna, D., Elhan, A. H., & Tennant, A. (2013). The reliability and validity of the World Health Organization Disability Assessment Schedule (WHODAS-II) in stroke. *Disability and Rehabilitation*, *35*(3), 214–220.

Kulnik, S. T., & Nikoletou, D. (2013). WHODAS 2.0 in community rehabilitation: a qualitative investigation into the validity of a generic patient-reported measure of disability. *Disability & Rehabilitation*, *36*(2), 146–154.

Larrabee, G. J. (2012). Performance Validity and Symptom Validity in Neuropsychological Assessment. *Journal of the International Neuropsychological Society*, *18*(04), 625–630. https://doi.org/10.1017/S1355617712000240

Lawton, G., Mayor, R. J., Howlett, S., & Reuber, M. (2009). Psychogenic nonepileptic seizures and health-related quality of life: The relationship with psychological distress and other physical symptoms. *Epilepsy & Behavior*, *14*(1), 167–171. https://doi.org/10.1016/j.yebeh.2008.09.029

Lee, G. P. (2010). *Neuropsychology of epilepsy and epilepsy surgery*. Oxford University Press.

Lezak, M. D. (1995). *Neuropsychological Assessment*. Oxford University Press: New York.

Lippa, S. M., Pastorek, N. J., Romesser, J., Linck, J., Sim, A. H., Wisdom, N. M., & Miller, B. I. (2014). Ecological validity of performance validity testing. *Archives of Clinical Neuropsychology*, *29*(3), 236–244.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.2307/2290157

Locke, D. E. C., Kirlin, K. A., Thomas, M. L., Osborne, D., Hurst, D. F., Drazkowski, J. F., … Noe, K. H. (2010). The Minnesota Multiphasic Personality Inventory-2-

Restructured Form in the epilepsy monitoring unit. *Epilepsy & Behavior*, *17*(2), 252–258. https://doi.org/10.1016/j.yebeh.2009.12.004

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

Mardia, K. V. (1980). 9 Tests of unvariate and multivariate normality. *Handbook of Statistics*, *1*, 279–320.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *General Hospital Psychiatry*, *28*(1), 71–77. https://doi.org/10.1016/j.genhosppsych.2005.07.003

Matsuoka, H. (2001). Neuropsychology of Epilepsy. *Epilepsia*, *42*, 42–46. https://doi.org/10.1046/j.1528-1157.42.s6.8.x

Mayeux, R., Brandt, J., Rosen, J., & Benson, D. F. (1980). Interictal memory and language impairment in temporal lobe epilepsy. *Neurology*, *30*(2), 120–120.

McKibbin, C., Patterson, T. L., & Jeste, D. V. (2004). Assessing disability in older patients with schizophrenia: results from the WHODAS-II. *The Journal of Nervous and Mental Disease*, *192*(6), 405–413.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal Of Clinical And Experimental Neuropsychology*, *24*(8), 1094–1102.

Muthén, L. K., & Muthén, B. O. (2002). How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Naismith, S. L., Longley, W. A., Scott, E. M., & Hickie, I. B. (2007). Disability in major depression related to self-rated and objectively-measured cognitive deficits: a preliminary study. *BMC Psychiatry*, *7*(1), 1–7. https://doi.org/10.1186/1471-244X-7-32

Ogunrin, O., Adamolekun, B., Ogunniyi, A., & Aldenkamp, A. (2000). Cognitive function in Nigerians with newly diagnosed epilepsy. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, *27*(2), 148–151.

Penny, K. I. (1996). Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance. *Applied Statistics*, 73–81.

Perini, S. J., Slade, T., & Andrews, G. (2006). Generic effectiveness measures: Sensitivity to symptom change in anxiety disorders. *Journal of Affective Disorders*, *90*(2–3), 123–130. https://doi.org/10.1016/j.jad.2005.10.011

Pösl, M., Cieza, A., & Stucki, G. (2007). Psychometric properties of the WHODASII in rehabilitation patients. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, *16*(9), 1521–1531. https://doi.org/10.1007/s11136-007-9259-4

Preacher, K. J., & Coffman, D. L. (2006). *Computing power and minimum sample size for RMSEA*. Retrieved from http://quantpsy.org/

Pulliainen, V., Kuikka, P., & Jokelainen, M. (2000). Motor and cognitive functions in newly diagnosed adult seizure patients before antiepileptic medication. *Acta Neurologica Scandinavica*, *101*(2), 73–78.

R Core Team. (2017). *R: A Language and Environment for Statistical Computing* [Windows]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271–276.

Rempfer, M. V., Hamera, E. K., Brown, C. E., & Cromwell, R. L. (2003). The relations between cognition and the independent living skill of shopping in people with schizophrenia. *Psychiatry Research*, *117*(2), 103–112.

Roeschl-Heils, A., Bledowski, C., Elger, C., Heils, A., & Helmstaedter, C. (2002). Neuropsychological functioning among 32 patients with temporal lobe epilepsy and their discordant siblings. *Epilepsia*, *43*(suppl 7), 85.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software. *Journal of Statistical Software*, *48*(2), 1–36.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Sellbom, M., Bagby, R. M., Kushner, S., Quilty, L. C., & Ayearst, L. E. (2011). Diagnostic Construct Validity of MMPI-2 Restructured Form (MMPI-2-RF) Scale Scores. *Assessment*, *19*(2), 176–186. https://doi.org/10.1177/1073191111428763

Sellbom, M., Ben-Porath, Y. S., & Bagby, R. M. (2008). On the hierarchical structure of mood and anxiety disorders: confirmatory evidence and elaboration of a model of temperament markers. *Journal of Abnormal Psychology*, *117*(3), 576.

Sellbom, M., Wygant, D., & Bagby, M. (2012). Utility of the MMPI-2-RF in detecting non-credible somatic complaints. *Psychiatry Research*, *197*(3), 295–301.

Shieh, G. (2010). Sample size determination for confidence intervals of interaction effects in moderated multiple regression with continuous predictor and moderator variables. *Behavior Research Methods*, *42*(3), 824–835. https://doi.org/10.3758/BRM.42.3.824

Smeets, V. M., van Lierop, B. A., Vanhoutvin, J. P., Aldenkamp, A. P., & Nijhuis, F. J. (2007). Epilepsy and employment: literature review. *Epilepsy & Behavior*, *10*(3), 354–362.

Spitzer, R. L., Williams, J. B., Kroenke, K., Linzer, M., Verloin deGruy, F., Hahn, S. R., … Johnson, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *Jama*, *272*(22), 1749–1756.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors (Vol. 758, pp. 424–453). Presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the beck depression inventory—second edition in a sample of college students. *Depression and Anxiety*, *19*(3), 187–189. https://doi.org/10.1002/da.20002

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics*. (S. J. Osterlind, Ed.) (6th ed.). Pearson.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.

Twamley, E. W., Doshi, R. R., Nayak, G. V., Palmer, B. W., Golshan, S., Heaton, R. K., … Jeste, D. V. (2002). Generalized Cognitive Impairments, Ability to Perform Everyday Tasks, and Level of Independence in Community Living Situations of

Older Patients With Psychosis. *American Journal of Psychiatry*, *159*(12), 2013–2020. https://doi.org/10.1176/appi.ajp.159.12.2013

Üstün, T. B. (2010). *Measuring health and disability: manual for WHO disability assessment schedule WHODAS 2.0*. World Health Organization.

Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., … Pull, C. (2010). Developing the World Health Organization disability assessment schedule 2.0. *Bulletin of the World Health Organization*, *88*(11), 815–823.

Van Rijckevorsel, K. (2006). Cognitive problems related to epilepsy syndromes, especially malignant epilepsies. *Seizure*, *15*(4), 227–234.

Ventura, J., Hellemann, G. S., Thames, A. D., Koellner, V., & Nuechterlein, K. H. (2009). Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: A meta-analysis. *Schizophrenia Research*, *113*(2–3), 189–199. https://doi.org/10.1016/j.schres.2009.03.035

Walczak, T. S., Papacostas, S., Williams, D. T., Scheuer, M. L., Lebowitz, N., & Notarfrancesco, A. (1995). Outcome After Diagnosis of Psychogenic Nonepileptic Seizures. *Epilepsia*, *36*(11), 1131–1137. https://doi.org/10.1111/j.1528-1157.1995.tb00472.x

Weber, M., Mapstone, M., Staskiewicz, J., & Maki, P. M. (2012). Reconciling subjective memory complaints with objective memory performance in the menopausal transition. *Menopause (New York, N.Y.)*, *19*(7), 735–741. https://doi.org/10.1097/gme.0b013e318241fd22

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*.

Wechsler, D. (2009). Wechsler Memory Scale-(WMS-IV). *New York: The Psychological Corporation.*

WHOQOL Group. (1993). Measuring quality of life: the development of the World Health Organization Quality of Life Instrument (WHOQOL). *Geneva: WHO*.

Wilkus, R. J., Dodrill, C. B., & Thompson, P. M. (1984). Intensive EEG monitoring and psychological studies of patients with pseudoepileptic seizures. *Epilepsia*, *25*(1), 100–107.

Williamson, D., Drane, D., Stroup, E., Miller, J., Holmes, M., & Wilensky, A. (2004). Detecting cognitive differences between patients with epilepsy and patients with psychogenic nonepileptic seizures: Effort matters (Vol. 45, pp. 179–179). Presented at the Epilepsia, BLACKWELL PUBLISHING INC 350 MAIN ST, MALDEN, MA 02148 USA.

Williamson, D. J., Drane, D. L., & Stroup, E. S. (2007). Symptom validity tests in the epilepsy clinic. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 346–365). Guilford Press.

Willment, K., Hill, M., Baslet, G., & Loring, D. W. (2015). Cognitive Impairment and Evaluation in Psychogenic Nonepileptic Seizures An Integrated Cognitive-Emotional Approach. *Clinical EEG and Neuroscience*, *46*(1), 42–53.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237

World Health Organization. (1980). International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29. 35 of the Twenty-ninth World Health Assembly, May 1976.

World Health Organization. (1988). WHO Psychiatric Disability Assessment Schedule (WHO/DAS) with a guide to its use. *Geneva: World Health Organization*.

World Health Organization. (2001). *International classification of functioning, disability and health: ICF*. World Health Organization.

Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 Restructured Form (MMPI-2-RF) Validity Scales in Civil Forensic Settings: Findings from Simulation and Known Group Samples. *Archives of Clinical Neuropsychology*, *24*(7), 671–680. https://doi.org/10.1093/arclin/acp073

Yuan, K.-H., & Bentler, P. M. (2000). Three Likelihood-Based Methods For Mean and Covariance Structure Analysis With Nonnormal Missing Data. *Sociological Methodology*, *30*(1), 165–200. https://doi.org/10.1111/0081-1750.00078

Zeber, J. E., Copeland, L. A., Amuan, M., Cramer, J. A., & Pugh, M. J. V. (2007). The role of comorbid psychiatric conditions in health status in epilepsy. *Epilepsy & Behavior*, *10*(4), 539–546.

Appendix 1. WHODAS 2.0, 36-item self-report version

WHODAS 2.0

World Health Organization Disability Assessment Schedule 2.0

36-item version, self-administered

Patient Name: _____ **Age:** _____ **Sex:** ❑ Male ❑ Female **Date:** _____

This questionnaire asks about difficulties due to health/mental health conditions. Health conditions include **diseases or illnesses, other health problems that may be short or long lasting, injuries, mental or emotional problems, and problems with alcohol or drugs.** Think back over the **past 30 days** and answer these questions thinking about how much difficulty you had doing the following activities. For each question, please circle only **one** response.

							Clinician Use Only		
Numeric scores assigned to each of the items:		1	2	3	4	5	Raw Item Score	Raw Domain Score	Average Domain Score
In the last 30 days, how much difficulty did you have in:									
Understanding and communicating									
D1.1	Concentrating on doing something for ten minutes?	None	Mild	Moderate	Severe	Extreme or cannot do			
D1.2	Remembering to do important things?	None	Mild	Moderate	Severe	Extreme or cannot do			
D1.3	Analyzing and finding solutions to problems in day-to-day life?	None	Mild	Moderate	Severe	Extreme or cannot do			
D1.4	Learning a new task, for example, learning how to get to a new place?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 30	___ 5
D1.5	Generally understanding what people say?	None	Mild	Moderate	Severe	Extreme or cannot do			
D1.6	Starting and maintaining a conversation?	None	Mild	Moderate	Severe	Extreme or cannot do			
Getting around									
D2.1	Standing for long periods, such as 30 minutes?	None	Mild	Moderate	Severe	Extreme or cannot do			
D2.2	Standing up from sitting down?	None	Mild	Moderate	Severe	Extreme or cannot do			
D2.3	Moving around inside your home?	None	Mild	Moderate	Severe	Extreme or cannot do			
D2.4	Getting out of your home?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 25	___ 5
D2.5	Walking a long distance, such as a kilometer (or equivalent)?	None	Mild	Moderate	Severe	Extreme or cannot do			
Self-care									
D3.1	Washing your whole body?	None	Mild	Moderate	Severe	Extreme or cannot do			
D3.2	Getting dressed?	None	Mild	Moderate	Severe	Extreme or cannot do			
D3.3	Eating?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 20	___ 5
D3.4	Staying by yourself for a few days?	None	Mild	Moderate	Severe	Extreme or cannot do			
Getting along with people									
D4.1	Dealing with people you do not know?	None	Mild	Moderate	Severe	Extreme or cannot do			
D4.2	Maintaining a friendship?	None	Mild	Moderate	Severe	Extreme or cannot do			
D4.3	Getting along with people who are close to you?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 25	___ 5
D4.4	Making new friends?	None	Mild	Moderate	Severe	Extreme or cannot do			
D4.5	Sexual activities?	None	Mild	Moderate	Severe	Extreme or cannot do			

							Clinician Use Only		
Numeric scores assigned to each of the items:		1	2	3	4	5	Raw Item Score	Raw Domain Score	Average Domain Score
In the last 30 days, how much difficulty did you have in:									
Life activities—Household									
D5.1	Taking care of your household responsibilities?	None	Mild	Moderate	Severe	Extreme or cannot do			
D5.2	Doing most important household tasks well?	None	Mild	Moderate	Severe	Extreme or cannot do			
D5.3	Getting all of the household work done that you needed to do?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 20	___ 5
D5.4	Getting your household work done as quickly as needed?	None	Mild	Moderate	Severe	Extreme or cannot do			
Life activities—School/Work									
If you work (paid, non-paid, self-employed) or go to school, complete questions D5.5–D5.8, below. Otherwise, skip to D6.1.									
Because of your health condition, in the past 30 days, how much difficulty did you have in:									
D5.5	Your day-to-day work/school?	None	Mild	Moderate	Severe	Extreme or cannot do			
D5.6	Doing your most important work/school tasks well?	None	Mild	Moderate	Severe	Extreme or cannot do			
D5.7	Getting all of the work done that you need to do?	None	Mild	Moderate	Severe	Extreme or cannot do		___ 20	___ 5
D5.8	Getting your work done as quickly as needed?	None	Mild	Moderate	Severe	Extreme or cannot do			
Participation in society									
In the past 30 days:									
D6.1	How much of a problem did you have in joining in community activities (for example, festivities, religious, or other activities) in the same way as anyone else can?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.2	How much of a problem did you have because of barriers or hindrances around you?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.3	How much of a problem did you have living with dignity because of the attitudes and actions of others?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.4	How much time did you spend on your health condition or its consequences?	None	Some	Moderate	A Lot	Extreme or cannot do		___ 40	___ 5
D6.5	How much have you been emotionally affected by your health condition?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.6	How much has your health been a drain on the financial resources of you or your family?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.7	How much of a problem did your family have because of your health problems?	None	Mild	Moderate	Severe	Extreme or cannot do			
D6.8	How much of a problem did you have in doing things by yourself for relaxation or pleasure?	None	Mild	Moderate	Severe	Extreme or cannot do			
						General Disability Score (Total):		___ 180	___ 5

Appendix 2. Selected Published WHODAS 2.0 Means Table

Published WHODAS 2.0 Total Scores in Various Clinical Samples

Sample Condition	WHODAS 2.0 Mean*	*N*	Source
Inflammatory arthritis	19.8	172	Baron et al., 2008
Epilepsy (pre-surgical)	16.7	22	Cankurtaran et al., 2005
Depression	27.14	73	Chwastiak & Von Korff, 2003
Back pain	22.75	76	
Healthy controls	12.95	271	Federici, Meloni, Mancini, et al., 2009
Motor disabled	28.66	111	
Mental disabled	24.60	45	
Sensory disabled	14.97	73	
Various chronic diseases[ǂ]	24.80	1190	Garin et al., 2010
Systemic Sclerosis	24.6	402	Hudson et al., 2008
Stroke	48.8	188	Küçükdeveci et al., 2013
Musculoskeletal conditions	22.0	296	Pösl et al., 2007
Internal conditions	18.5	308	
Stroke	38.7	116	
Breast cancer	23.8	119	
Depression	44.6	65	
Mental health problems[§]	33	-	Üstün et al., 2010
Alcohol problems[§]	22.5	-	
Drug problems[§]	31	-	
Physical health problems[§]	29.5	-	
General population[§]	6	-	

Note. WHODAS 2.0, World Health Organization Disability Assessment Schedule 2.0.
*Means are reported to significant digits in the original publication.
[ǂ]Total score not reported for diagnostic subgroups.
[§]Values estimated from published chart. Means refer to values obtained in validation studies; exact values not reported in publication.