University of Windsor
Scholarship at UWindsor

Electronic Theses and Dissertations | Theses, Dissertations, and Major Papers

9-26-2018

Improving Diagnostic Accuracy in HIV-Associated Neurocognitive Disorders

Jessica Hurtubise
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
Hurtubise, Jessica, "Improving Diagnostic Accuracy in HIV-Associated Neurocognitive Disorders" (2018). *Electronic Theses and Dissertations*. 7527.
https://scholar.uwindsor.ca/etd/7527

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

Improving Diagnostic Accuracy in HIV-Associated Neurocognitive Disorders

By

Jessica Hurtubise, MSc.

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Master of Arts
at the University of Windsor

Windsor, Ontario, Canada

2018

© 2018 Jessica Hurtubise

Improving Diagnostic Accuracy in HIV-Associated Neurocognitive Disorders

by

Jessica Hurtubise

APPROVED BY:

_____

E. Maticka-Tyndale
Department of Sociology, Anthropology, and Criminology

_____

C. Abeare
Department of Psychology

_____

L. Erdodi, Advisor
Department of Psychology

July 12, 2018

DECLARATION OF CO-AUTHORSHIP

I hereby declare that this thesis incorporates material that is result of joint research, as follows:

Part of Chapter 2 was co-authored with Dr. Laszlo Erdodi. The key ideas, primary contributions, experimental designs, data analysis, interpretation, and writing were performed by the author, and the contribution of co-authors was primarily through the provision of guidance and supervision. Both Dr. Abeare and Dr. Maticka-Tyndale provided feedback on the refinement of ideas and editing of the manuscript.

The archival dataset was collected from the Southern Alberta Clinic in Calgary, Alberta by Daniella Gomez, Dr. Fujiwara, Dr. Power, and Dr. Gill.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-authors to include the above material in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada

Copyright Act, I certify that I have obtained a written permission from the copyright owner to include such material in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Human immunodeficiency virus (HIV) associated neurocognitive disorders (HAND) affect 50% of individuals with HIV. HAND is characterized by cognitive and functional impairment and is diagnosed through neuropsychological assessment. The use of performance validity tests (PVT) is recommended to determine the credibility of cognitive profiles during neuropsychological testing. However, little is known about the utility of PVTs within an HIV+ population. The objective of the present study was to compare the base rate of failure on embedded validity indicators (EVIs) between individuals diagnosed with HAND, neurocognitively normal individuals with HIV, undergraduate controls, and undergraduates asked to feign cognitive impairment. The relationship between EVI failure and neurocognitive performance, as well as self-reported depressive symptoms, was also explored. Cumulative EVI failure produced good classification accuracy within the student sample, reaffirming their utility in detecting invalid performance. As predicted, individuals with more severe HAND diagnoses (i.e., HIV-associated dementia and mild cognitive impairment) failed more EVIs than neurocognitively normal individuals. Further, as neurocognitive test performance decreased, cumulative EVI failures increased. Although directionality of this finding could not be determined (i.e., do low scores reflect non-credible responding or are EVI failures false positives in individuals with genuine impairment?), monitoring performance validity might help explain the well-known fluctuation in cognitive performance over time in the HAND population. There was no relationship between the number of EVIs failed and self-reported depressive symptoms or severity, ruling out a commonly discussed confounding variable in PVT research.

ACKNOWLEDGEMENTS

I would like to thank all the individuals at the Southern Alberta Clinic who agreed to have their neurocognitive data used for research purposes. I hope this data will be used to help provide you with the most appropriate and optimal treatment. I would also like to thank all the undergraduate students who devoted their time to the present study.

Thank you to Dr. Erdodi for supporting me throughout the conceptualization, execution, and writeup of this thesis project. Thank you for compromising with me, sharing your knowledge, and encouraging my work-life balance. Additionally, thank you for your patience and time during incredibly tight timelines.

This project would not have been possible without the support of Dr. Fujiwara, Dr. Power, and Dr. Gill. They were willing to share data collected over multiple years for this project and I hope it provides a platform to begin measuring performance validity in HIV research.

I would like to thank my committee members, Dr. Christopher Abeare and Dr. Eleanor Maticka-Tyndale for their enthusiasm about my project, asking difficult questions, and their constructive criticism.

I don't think I would have been able to finish the past two years without my incredible social network – my family and friends. Specifically, late nights with Tara McAuley and Brandon Zuccato provided so much support and joy during difficult times. I am also very grateful for my friends that have offered laughter, guidance, and relief including Amy Tran, Sanya Sagar, John Howland, and Christie Kimber. Lastly, I would like to thank my family for their unconditional love.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
ANI	Asymptomatic Neurocognitive Impairment
ANOVA	Analysis of Variance
AUC	Area Under the Curve
BR_{Fail}	Base Rate of Failure
cART	Combination Antiretroviral Therapy
CD4	Cluster of Differentiation 4
CNS	Central Nervous System
CONT	Undergraduate Controls
EXP_{MAL}	Experimental Malingering Condition
EVI	Embedded Validity Indicator
FMS	Failure to Maintain Set
GPB	Grooved Pegboard Test
HAD	HIV-Associated Dementia
HAND	HIV-Associated Neurocognitive Disorder
HDS	HIV Dementia Scale
HIV	Human Immunodeficiency Virus
HIV+	Individuals Living with HIV
HDS	HIV Dementia Scale
HVLT-R	Hopkins Verbal Learning Test – Revised
MND	HIV-Associated Mild Neurocognitive Disorder
NN	Neurocognitively Normal Individuals Diagnosed with HIV

PHQ-9	Patient Health Questionnaire
PVT	Performance Validity Test
SAC	Southern Alberta Clinic
SD	Standard Deviation
SDMT	Symbol Digit Modalities Test
TBI	Traumatic Brain Injury
TMT	Trail Making Test
WCST	Wisconsin Card Sorting Test
WCT	Word Choice Test
WRAT-4	Wide Range Achievement Test (4th Version)

Chapter I

Introduction

Performance Validity Testing

In neuropsychological assessment, accurate diagnosis and appropriate treatment are dependent upon the valid presentation of an examinee's neuropsychological functioning. Often lacking objective biomarkers, a neuropsychologist's decisions are based on a combination of self-reported symptoms, behavioural observations, and performance on objective tests of cognitive ability (Carone, 2015). However, there is a growing awareness that neuropsychological test performance is not always an accurate reflection of an examinee's true ability. Performance validity tests (PVTs) were developed to assess whether (or the extent to which) the scores on neuropsychological tests are an accurate reflection of the examinee's neurocognitive functioning (Boone, 2013; Larrabee, 2014b).

PVT failures are commonly interpreted as evidence of non-credible responding and alert the assessor that test results may be invalid and should be interpreted with caution. Non-credible performance, as indicated by PVT failure, may explain up to 50% of variance on neuropsychological testing (Green, Rohling, Lees-Haley, & Allen, 2001; Meyers, Volbrecht, Axelrod, & Reinsch-Boothby, 2011), has produced large effect sizes (d = 1.0), and ultimately diminished study replicability (Larrabee, 2012).

Brief history. In 1912, Sir John Collie discussed malingering as a significant concern within medical practice (as referenced in Greher & Wodushek, 2017). This marked the first recorded instance of validity concerns within a health care setting. By the 1940's, Andre Rey had developed the Rey-15 item and dot counting tests. These were the first indicators of performance validity in neuropsychological testing (Frederick, 2003).

In the 1990's, PVT development rapidly increased (Carone, 2015). This growth was a consequence of the gradual realization that the clinical judgment of psychologists and psychiatrists about malingering during expert witness testimonies were often unreliable and inaccurate (Faust, Hart, Guilmette, & Arkes, 1988a, 1988b; Heaton, Smith, Lehman, & Vogt, 1978). In addition, it was during this time period that the first modern book on malingering was published (Rogers, 1988). Validity testing has since extended beyond neuropsychological testing. Objective measures of the under- and over-reporting of symptoms have been incorporated into multiple psychological inventories and structured interviews, such as the L and K scales in the Minnesota Multiphasic Personality Inventory (Ben-Porath & Tellegen, 2008) as well as the infrequent and negative impressions subscales of the Personality Assessment Inventory (Morey, 2007).

Recent literature suggests that the majority of assessors have incorporated measures of performance validity into their practices (Dandachi-FitzGerald, Merckelbach, & Ponds, 2017; Jung & Reidenberg, 2007). In 2015, over 92% of 316 neuropsychologists surveyed reported "often" or "always" using a PVT to detect non-credible performance (Martin, Schroeder, & Odland, 2015). This is a dramatic increase from a 2007 study reporting only 52% of neuropsychologists frequently used PVTs (Sharland & Gfeller, 2007). The importance of PVTs in clinical practice has been highlighted by several professional organizations. The National Academy of Neuropsychology and the American Academy of Clinical Neuropsychology have declared that PVTs are "medically necessary" and "important in all evaluations" (Board of Directors, 2007; Bush et al., 2005).

Invalid performance. Concerns of non-credible presentation vary across clinical populations and settings. Base rate of failure (BR_{Fail}) indicates the proportion of individuals

within a population who fail a PVT. BR_{Fail} on validity tests during neuropsychological exams are highest for cases involving personal injury, workers compensation, criminal justice, and diseases lacking clear neuropathological biomarkers (Mittenberg, Patton, Canyock, & Condit, 2002). For example, chronic pain patients with financial incentive to perform poorly and individuals in criminal forensic settings have non-credible performance rates as high as 50% (Ardolf, Denney, & Houston, 2007; Greve, Ord, Bianchini, & Curtis, 2009). In psychoeducational evaluations of ADHD where external incentives such as medication and academic accommodations are present, 25-50% of examinees are believed to exaggerate their deficits (Marshall et al., 2010; Suhr, Hammers, Dobbinsbuckland, Zimak, & Hughes, 2008; Sullivan, May, & Galbally, 2007).

Motivation to exaggerate or feign deficits is not the only causal mechanism behind invalid performance. In settings without identifiable external incentives or in diseases with objective biomarkers, BR_{Fail} are estimated around 10% (Mittenberg et al., 2002). Other explanations may include emotional distress, somatic concerns, fatigue, pain, sensory disturbances, and limited English proficiency (Erdodi, Seke, et al., 2017; Erdodi, Nussbaum, Sagar, Abeare, & Schwartz, 2017; Erdodi et al., 2016; Greher & Wodushek, 2017, Whiteside et al., 2010). Other psychological diagnoses that can increase the risk of PVT failure include factitious disorder, oppositional behaviour, and personality disorders (Carone, 2015).

It is worth noting that invalid performance and genuine impairment are not mutually exclusive. In child custody cases, up to 98.3% of parents pass PVTs (Flaro, Green, & Robertson, 2007). In other words, external incentive to appear cognitively intact on neuropsychological tests dramatically reduces BR_{Fail}. Interestingly, in situations without external incentive to appear cognitively intact or impaired, BR_{Fail} remains relatively high. For example, the cognitive functioning of undergraduate students is typically higher than average. However, while

completing neuropsychological tests for research purposes, 37% failed at least 1 PVT (An, Kaploun, Erdodi, & Abeare, 2017). Thus, it appears that the absence of an apparent external incentive to perform poorly is not equivalent to the presence of incentives to do well.

Terminology. The language surrounding validity tests has evolved since their initial conception (Greher & Wodushek, 2017). Originally, PVTs were believed to detect malingering (Slick, Sherman, & Iverman, 1999). Malingering is the fabrication or intentional exaggeration of symptoms motivated by secondary external incentives (American Psychiatric Association, 2013). Based on criteria proposed by Slick et al. (1999), malingering could be identified and further classified according to the level of confidence associated with the diagnosis (e.g. probable, possible, or definite). It has since become clear that, although secondary gain and ligation may motivate an examinee to perform poorly, intent cannot be definitively known. In turn, PVT failure was gradually re-labeled as "poor effort" (Van Dyke, Millis, Axelrod, & Hanks, 2013). However, *effort* remains an ill-defined construct, and carries residual connotations of *intent* (i.e., "not trying hard enough"). Moreover, simulating poor effort produces activation peaks in the same cortical regions as full effort conditions on the Word Memory Test (Larsen, Allen, Bigler, Goodrich-Hunsaker, & Hopkins, 2010). Thus, when *effort* is quantified using a measure of neural activation, no differences are observed between those intentionally performing poorly and those trying their best. There may even be a unique pattern of neural activation that occurs while being deceptive (Kireev, Korotkov, Medvedeva, & Medvedev, 2013), further supporting the idea that suppressing true ability level while trying to avoid detection likely requires significant mental energy (i.e., effort). Since many neuropsychological tests begin with statements like "try your best" or "give your best effort", failing PVTs can be conceptualized as non-compliance with instruction (Slick & Sherman, 2013).

Since 2003, the preferred terminology used to describe PVT failure has been suboptimal, non-credible, or invalid performance (Boone & Lu, 2003). While still communicating that neuropsychological test scores may not reflect true cognitive ability, "non-credible" does not imply etiology (e.g. motivation or volition). Additionally, this language allows the conclusion to be objective and data-driven, while the clinician remains unbiased (Rickards, Cranston, Touradji, & Bechtold, 2017).

Free-standing vs. embedded. By design, there are two types of PVTs: free-standing (stand-alone) and embedded. Free-standing PVTs are independently administered and their primary purpose is to estimate the credibility of a response set. Although they appear to measure neurocognitive performance, free-standing PVTs are largely insensitive to brain function and give little-to-no insight into cognitive ability (Greher & Wodushek, 2017). Commonly used free-standing PVTs include the Rey-15 item Test (Rey, 1964), Test of Memory Malingering (TOMM; Tombaugh, 1996), and the Word Choice Test (WCT; Martin et al., 2015; Pearson, 2009). Forced choice recognition is the most commonly utilized paradigm for free-standing PVTs (Bigler, 2014). In a forced-choice recognition task, an examinee is presented with a set of stimuli. Later, when target and foil(s) are presented, the examinee is instructed to identify the previously presented stimulus (Pankratz, 1983). However, one significant limitation of free-standing PVTs is their extension of overall assessment time without contributing any information regarding the examinee's current cognitive functioning (Rickards et al., 2017).

Unlike free-standing PVTs, embedded validity indicators (EVIs) are derived from traditional tests of cognitive ability and therefore, add no extra administration time. As such, they allow the simultaneous assessment of performance validity and neuropsychological functioning throughout the testing session. At sufficiently conservative cutoffs, EVIs are insensitive to

neurological diseases and are indicative of non-credible performance rather than cognitive deficit. EVIs are difficult to identify as PVTs, making them resistant to coaching and preserving their psychometric utility (Schutte & Axelrod, 2013). In addition, EVIs are advantageous in that they directly assess the credibility of a specific response set rather than inferring it through scores on PVTs administered at different times throughout the assessment (Suhr & Gunstad, 2000).

The mechanisms by which EVIs identify non-credible performance are versatile. The most common detection method relies on a demonstrated psychometric floor, beyond which a score is unlikely to occur (Greher & Wodushek, 2017). In other words, EVI failure may be construed as a deficit so severe that it is rarely observed in clinical populations, raising questions about its credibility (Babikian, Boone, Lu, & Arnold, 2006). Alternatively, EVIs may identify errors that are highly unusual even in cases of severe neurological dysfunction. An example of this is failure to maintain set (FMS) errors within the Wisconsin Card Sorting Task (Greve, Bianchini, Mathias, Houston, & Crouch, 2002; Lichtenstein, Erdodi, Rai, Mazur-Mosiewicz, & Flaro, 2016; Suhr & Boyer, 1999). Similarly, another type of EVI looks at atypical patterns across tests including performing better on more difficult tests compared to easier ones requiring the same basic skill. For example, an "atypical profile" occurs when an individual performs significantly better on a WAIS-R Vocabulary subtest compared to the Digit Span subtest (Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995) or when a large discrepancy is observed in the age-corrected scaled score on the Coding and Symbol Search subtests, two analogous measures of psychomotor processing speed (Erdodi, Abeare, et al., 2017; Glassmire, Wood, Ta, Kinney, & Nitch, 2018).

The strength of EVIs is also a source of liability in that they are closely linked to the cognitive function assessed by the test they are embedded within. Therefore, depending on the cutoff used, EVIs may be more prone to false-positives than free-standing tests (DeRight & Carone, 2015). One way to minimize the confound of cognitive ability on PVT outcome is to consider EVI's across cognitive domains and/or aggregate multiple EVIs into a single validity composite (Erdodi, Nussbaum, et al. 2017).

Cutoffs. Cutoff scores separate non-credible performance and genuine dysfunction (Slick et al., 1999). A score on the passing side of the cutoff represents valid performance whereas a score on the failing side of the cutoff is interpreted as evidence of invalid performance (Bigler, 2014). PVTs are optimized to minimize the rate of false positives (i.e., maximize specificity) at the expense of sensitivity. A generally accepted specificity rate is ≥.90, resulting in a less than 10% false positive rate (Larrabee, 2014b).

Cutoff scores are determined *a priori*, based on previous literature, and allow clinicians to assess performance validity and estimate the likelihood that the profile is invalid. When applying cutoffs developed on a given clinical populations to a different diagnostic group it is important that clinicians reflect on the implications on classification accuracy. For example, an inflated false positive rate was reported when used on reliable digit span cutoffs in populations with severe memory disorders, cerebrovascular accidents, and children (Blaskewitz, Merten, & Kathmann, 2008; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012). Certain populations, such as dementia and intellectual disability, are exempt from PVTs due to the combination of well-established severe neurological impairment and high BR_{Fail}. Given the lack of universally applicable cutoffs, selecting a cutoff for a population that is yet to be validated carries the risk of diminished classification accuracy.

Minimizing false positives. The use and interpretation of cutoff scores varies across settings, populations, and assessors. However, there is a consensus that PVTs should be optimized for specificity (i.e., to minimize false positive rates). Larrabee (2014b) identified six cognitive domains that should be assessed within a comprehensive clinical neuropsychology battery: (1) verbal symbolic abilities; (2) visuoperceptual and visualspatial judgement and problem solving; (3) sensorimotor function; (4) attention/working memory; (5) processing speed; (6) learning and memory-verbal and learning and memory-visual. While most free-standing PVTs are memory based, EVIs have been developed within each of the six domains. Larrabee (2014b) recommends completing assessments using neuropsychological batteries that include embedded measures from all 6 cognitive domains as well as free-standing PVTs. His sample battery contains a total of 27 tests, 10 of which include EVIs, and recommends additional free-standing PVTs.

Careful consideration must be taken when deciding how many PVTs to use, how to minimize the burden of additional PVTs, and which cutoff scores are appropriate for each particular examinee (Rickards, Cranston, Touradji, & Bechtold, 2017). Although it is generally agreed that a single PVT failure provides insufficient evidence to determine non-credible performance, the exact number of PVT failures required to deem an entire neurocognitive profile invalid varies across assessors. It has been argued that false positive risk increases substantially with the number of PVTs given (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013; Bilder, Sugar, & Hellemann, 2014), whereas many contend the risk of false positives can be reduced by responsibly adjusting cutoff scores or the number of PVT failures needed for the profile to be considered invalid (Larrabee, 2014a; Odland, Lammy, Martin, Grote, & Mittenberg, 2015). Using multiple PVTs can increase sensitivity without reducing specificity because the

probability of having multiple PVT failures is low (Jasinski et al., 2011; Victor, Boone, Serpa, Buehler, & Ziegler, 2009). In fact, Larrabee (2014a) argued that multivariate models of performance validity assessment can protect against false positive errors.

A final consideration when determining the likelihood of invalid performance may be the level of failure. This is particularly true in a forced-choice paradigm. For example, if an individual answers 15% of the items correctly, a performance well-below chance level responding on the TOMM, the score can be confidently interpreted as non-credible (Slick & Sherman, 2012). Rickards, Cranston, Touradji, and Bechtold (2017) offered a systematic approach to PVT administration by creating a decision tree for neuropsychologists to follow when deciding whether or not to administer additional PVTs to determine credibility. The authors suggest that several factors should be considered when evaluating performance credibility including identification of risk groups (incentive, referral type, patient population), behavioural observations (non-credible symptom endorsement, inconsistencies in self-reports), and PVT and symptom validity outcomes.

In summary, in order to minimize the likelihood of a false positive error, neuropsychologists should utilize multiple independent tests (both embedded and free-standing) with high sensitivity and specificity that cover a variety of cognitive domains. Tests should be interpreted in the context of the overall evaluation and assessors should ensure that the PVT cutoffs used are appropriate for the population. To the best of our knowledge, no research has been conducted on PVT use in a population with HIV-associated neurocognitive disorders (HAND).

HIV-Associated Neurocognitive Disorders

HAND is cognitive dysfunction secondary to human immunodeficiency virus (HIV) infection. As many as 50% of HIV-positive (HIV+) individuals have neurocognitive impairment, which is associated with unemployment and reduced independence in daily living (Heaton, Marcotte, et al., 2004, 2010). Further, neurocognitive dysfunction in individuals with HIV increases mortality risk (Ellis et al., 1997).

Frascati criteria. In 1991, the American Academy of Neurology proposed the diagnosis of HAND be divided into two subtypes based upon the severity of cognitive and daily living impairments (Janssen et al., 1991). However, this subdivision was imprecise and insufficient as it did not specify criteria (i.e. the extent of impairment) nor did it allow for the diagnosis of patients with cognitive but not functional impairments. To address these shortcomings, the HIV Neurobehavioural Research Center proposed a new way to categorize the HAND diagnosis, now known as the Frascati criteria. This group identified three distinct categories: HIV-associated dementia (HAD), HIV-associated mild neurocognitive disorder (MND), and asymptomatic neurocognitive impairment (ANI; Antinori et al., 2007). To be classified into any of the three categories, differential diagnoses and comorbidities must be ruled out as the principal etiology for neurocognitive deficits.

HIV-associated dementia. Individuals with severe cognitive and functional impairments are diagnosed with HAD. To be placed in this category, patients must perform ≥2 standard deviations (SD) below the normative mean on two neurocognitive domains. They must also indicate moderate-to-severe levels of impairment in daily living, as assessed by self- or informant- reports, or be impaired on standardized measures of activities of daily living (Antinori et al., 2007). Although the number of HAND cases persists (Heaton et al., 2010), the incidence

of HAD has declined since the introduction of combination antiretroviral therapy (cART) to HIV treatment (Sacktor et al., 2002).

The cognitive impairment in HAD has been associated with reduced glutamate and glutamine levels within the frontal white matter, which may be indicative of neuronal dysfunction (Mohamed et al., 2010). Patients with HAD perform poorly on verbal fluency tasks, producing a greater number of errors and fewer total words than a non-demented HIV+ sample (Woods, 2004). On auditory learning and memory tests, words from the end of the list are overrepresented in HAD patients. This pronounced recency effect is commonly interpreted as an emergent sign of severe memory deficit, as the examinee is compensating for impaired encoding and consolidation skills by increasingly relying on auditory attention/echoic memory – cognitive domains that are relatively robust to the deleterious effects of neurodegenerative diseases (Scott et al., 2006).

Mild neurocognitive disorder. Individuals whose cognitive deficits lead to mild impairments in daily living are categorized as having MND. In order to meet the recommended criteria, a patient must score ≥1 SD below the normative mean in at least two cognitive domains. Mildly impaired daily living includes reports of diminished independence, accuracy, or efficiency in adaptive or occupational functioning, operationalized as a score ≥1 SD below normative means on standardized functional tests, and/or informant-/ self- reports of requiring assistance in ≥2 cognitive domains or activities (Antinori et al., 2007).

Asymptomatic neurocognitive impairment. The ANI classification allows for the objective diagnosis of patients presenting with cognitive dysfunction but no impairment in daily living. Similar to MND, the patient must score ≥1 SD below the normative mean in ≥2 cognitive domains but does not meet the remaining criteria of MND and HAD (Antinori et al., 2007). Over

half of those diagnosed with HAND only meet the criteria for ANI. This finding is consistent across geographic regions including Brazil (de Almeida et al., 2017), Korea (Ku et al., 2014), and China (Zhao et al., 2015).

Prognosis. The prognosis of HAND fluctuates over time and across cases. Over one year, only 58% of patients had stable cognitive performance, while the remaining 42% either worsened, improved, or oscillated between the two (Antinori et al., 2007). If a patient's cognitive performance improves to the point that their HAND categorization is no longer appropriate, the specifier "in remission" may be added. cARTs appear to increase performance in several cognitive domains and improve prognosis up to 4 years after treatment initiation (Kore et al., 2015; Robertson et al., 2012; Willen, Cuadra, Arheart, Post, & Govind, 2017).

Ultimately, cognitive changes are difficult to predict and a patient's neurocognitive ability should be continually monitored (Kamminga et al., 2017). Neurocognitive impairment has been associated with a variety of variables including lower cluster of differentiation 4 (CD4) counts, cART, cerebral spinal fluid HIV ribonucleic acid, smoking, neuropathy, substance use, stress, and body mass index (Akhtar-Khaleel et al., 2017; Chang, Lim, Lau, & Alicata, 2017; Cohen et al., 2011; Fellows et al., 2012; Gustafson et al., 2013; Keen & Turner, 2014; Kinuthia, Thigiti, & Gakinya, 2016; Kore et al., 2015; Muñoz-Moreno et al., 2008, 2013; Rubin et al., 2015). Further, cytokine levels within the cerebral spinal fluid have been associated with slowed psychomotor speed and impaired executive functioning (Nolting et al., 2012). More specifically, high interleukin 6 levels have been linked to reduced processing speed abilities as demonstrated by impairments on the Symbol-Digit Modalities Test (SDMT) and Trail Making Test (TMT; Keen & Turner, 2014; Lake et al., 2015).

Diagnosis. In order to categorize a HAND patient, a clinician must determine the presence and severity of their neurocognitive dysfunction. The 3 screening tools measuring neurocognitive deficits in HAND are traditional neuropsychological testing, the CogState computerized battery, and the demographically adjusted HIV dementia scale (HDS; Kamminga et al., 2017). However, only neuropsychological measures and the CogState battery have a sensitivity and specificity above .70 for detecting HAND (Cysique, Maruff, Darby, & Brew, 2006; de Almeida et al., 2017; Moore et al., 2012). Further, the CogState battery has been developed as a research tool and has not been validated for clinical use. Therefore, the gold standard for cognitive testing in HAND research and clinical assessment is a neuropsychological battery (de Almeida et al., 2017; Kamminga et al., 2017).

Some limitations of neuropsychological tests include their cost (clinician time, test material) and dependence on appropriate norms. When determining a patient's relative standing compared to a strategically selected comparison group, it is important that relevant demographic variables known to influence test performance (age, education, and gender) are accounted for. Equally important are ethnic and racial background, with one study finding 71% of HIV+ African Americans were considered cognitively impaired when using the Caucasian norms, but this number was reduced 45% when using African American norms (Antinori et al., 2007). Similarly, once psychosocial and environmental factors were accounted for, group differences in executive functioning between HIV+ and HIV- children diminished (Llorente et al., 2014). Taken together, these studies emphasize the role of demographic factors in neuropsychological test outcomes and highlight the need of appropriate norms.

Neurocognitive testing. Antinori and colleagues (2007) outlined 7 domains of interest in neuropsychological testing, specifying that at least 5 domains should be examined prior to

HAND classification: attention, language, executive functioning, motor skills, memory/learning, processing speed, and sensory-perceptual abilities. The authors further specify that at least one of the deficits must be cognitive in nature, eliminating the diagnosis if deficits are only observed in sensory-perceptual and motor areas. There are a variety of tests that could be used to examine the six domains but some of the most common ones include the Hopkins Verbal Learning Test – Revised (HVLT-R), Grooved Pegboard Test (GPB), SDMT, TMT, Wisconsin Card Sorting Test (WCST), and verbal fluency measures (de Almeida et al., 2017; Eggers et al., 2017; Gomez, Power, Gill, & Fujiwara, 2017; Zhao et al., 2015). Following the introduction of cARTs to HIV treatment, the primary pattern of cognitive impairment shifted from reduced psychomotor and cognitive speed to impaired memory and executive functioning (Heaton et al., 2011). Still, impairments occur in all domains of interest within this population and should be assessed using a comprehensive battery of neuropsychological tests.

Within the HIV+ population, verbal learning and memory are commonly measured using the HVLT-R. Women with HIV have significantly reduced performance on total learning and delayed recall (Spies, Fennema-Notestine, Archibald, Cherner, & Seedat, 2012; Woods et al., 2005). Interestingly, the HVLT-R has been identified as one of the most sensitive indicators of HAND and, as such, has been included in several brief screening batteries. When a two-test battery is used, the combination of HVLT-R total recall and non-dominant hand GPB T-score <40 had .78 sensitivity and .85 specificity (Carey et al., 2004). The combination of HVLT-R and the Stroop test T-score <40 on both or <35 on one of the tests had a sensitivity of .73 and a specificity of .83. (Moore et al., 2012).

The GPB is a measure of fine motor functioning and psychomotor speed. HIV-associated deficits in motor skills may relate to desynchronization between primary motor cortex and

supplementary motor areas (Wilson et al., 2013) and/or grey matter atrophy within the basal ganglia (Küper et al., 2011).

Within an HIV+ population, SDMT and the number-sequencing trial of the TMT (TMT-A; D-KEFS Trails 2) is used to measure processing speed. HIV+ women perform worse on the SDMT total correct compared to HIV- controls even after education, age, ethnicity, and reading level are accounted for (Manly et al., 2011). In an HIV+ East Indian cohort, impaired performance on the TMT and SDMT was unaffected by illness duration during the early stages of the disease (Mandal et al., 2008). Impairments on these tests are not as severe in patients with intact immunological functioning and suppressed viral load, suggesting that low scores may be a consequence of the breakdown of immune systems in HIV (Cole et al., 2007). Interestingly, physical activity appears to protect against the deleterious effects of HIV on neuropsychological tests measuring attention (Monroe et al., 2017).

Executive functions in an HIV+ population is typically measured using the WCST and the letter-number sequencing trial on the TMT (TMT-B; D-KEFS Trails-4). Compared to HIV-controls, performance on both tests is impaired in HIV+ patients, with the greatest impairments seen in those with acquired immunodeficiency syndrome (AIDS; Basso & Bornstein, 2003; Moradi, Miraghaei, Parhon, Jabbari, & Jobson, 2012). Further, compared to demographically matched controls, individuals perinatally infected with HIV have lower D-KEFS Trails-4 performance (Willen, Cuadra, Arheart, Post, & Govind, 2017). Performance on the TMT-B appears to decline much more rapidly over time in older adults with HIV than those without, suggesting that being HIV+ accelerates age-related cognitive decline (Sacktor et al., 2010). Executive functioning impairments on the WCST in an HIV+ sample have been correlated with a reduced caudate nucleus volume (Corrêa et al., 2016).

Finally, letter and category fluency are commonly used to measure verbal abilities in an HIV+ samples (Cysique et al., 2011). HIV+ status does not appear to dramatically impair performance on verbal fluency tests (Thames et al., 2016). A meta-analysis of available literature found the effect size for impairments due to HIV are small and similar between letter and category fluency tests (Iudicello et al., 2008). Still, HIV is associated with psychometrically detectable word generation deficits (Iudicello et al., 2007, 2008).

Performance validity. As mentioned, neuropsychological test performance can be influenced by a variety of factors outside an examinee's cognitive ability and several potential confounds may be present in an HIV+ population. Seventy-three percent of HIV+ patients have been classified as having a sleep disturbance according to the Pittsburgh Sleep Quality Index. Insomnia was particularly common among individuals with cognitive impairments (Rubinstein & Selwyn, 1998). The findings of a meta-analysis conducted in 2015 found that 58% of HIV+ patients had self-reported sleep disturbances. Across populations, North America had the greatest prevalence of sleep disturbances in this population, reaching over 70% (Wu, Wu, Lu, Guo, & Li, 2015). The North American HIV+ population also has high levels of self-reported pain (Lawson et al., 2014). A systematic review of 61 studies found the prevalence of pain ranged between 54% to 83% and was most commonly of moderate-to-severe intensity (Parker, Stein, & Jelsma, 2014). The relationship of PVTs with sleep and pain within the literature is inconsistent. For example, in traumatic brain injury (TBI) sleep appears to have no effect on performance validity (Dean & Sterr, 2013). However, PVT outcome has been shown to correlate with sleep and pain in individuals with fibromyalgia (Johnson-Greene, Brooks, & Ference, 2013). Further, PVT BR_{Fail} was as high as 50% in individuals with chronic regional pain syndrome type 1 in litigative settings (Greiffenstein, Gervais, Baker, Artiola, & Smith, 2013). Further, individuals with HIV

also report symptom and emotional distress (Jaggers et al., 2014; Pereira, Fialho, & Canavarro, 2014). Importantly, there is a high level of apathy in HIV+ patients (McIntosh, Rosselli, Uddin, & Antoni, 2015). Taken together, all of these comorbid symptoms associated with HIV may increase the likelihood of non-credible responding in HIV+ populations during neuropsychological testing.

Research aimed at understanding what the typical PVT profile looks like in an HIV+ population will facilitate the detection of non-credible response sets. Developing a psychometric method for differentiating valid and invalid profiles would allow clinicians and researchers to determine with greater confidence whether low scores on neuropsychological tests, and the corresponding HAND diagnosis, reflect true impairment or non-credible performance. As stated earlier, 42% of patients have changes in their neurocognitive performance over a 1-year time span (Antinori et al., 2007). Part of this may reflect fluctuations in performance validity. Therefore, research aimed at understanding performance validity in the HIV+ population has important psychometric implications that could improve diagnostic certainty and disease management.

Objective

The overarching objective of this study is to assess performance validity in an HIV+ sample. To achieve this objective, archival HIV+ patient performance was compared to prospectively-collected undergraduate scores on EVIs contained within a HAND battery. To the best of our knowledge, the present study is the first to explore PVTs in an HIV+ sample. There were four groups within the HIV+ sample, those identified as neurocognitively normal (NN) and those diagnosed with ANI, MND, or HAD. EVI scores were compared between these four groups as well as to an undergraduate sample. Undergraduate participants were assigned to either

a control condition, where they were asked to try their best, or an experimental malingering condition (EXP_{MAL}), where they were instructed to feign neurocognitive deficits. Therefore, a total of six groups existed within this study: NN, ANI, MND, HAD, controls, and EXP_{MAL}.

In the first part of the study, we compared the BR_{Fail} across groups – both on individual measures and the cumulative failure rates (i.e., number of individuals who failed ≥1, ≥2, ≥ 3, ≥ 4, and ≥5 PVTs). We hypothesized that EXP_{MAL} would have the highest BR_{Fail} as they are instructed to intentionally perform below their true ability. Additionally, we hypothesized that the level of EVI failure in EXP_{MAL} would not differ from the HAD sample because dementia samples are prone to highly elevated BR_{Fail} on measures of performance validity (Davis, 2018). The next highest BR_{Fail} was expected in the MND and ANI samples. As certain EVIs have been shown to be sensitive to genuine cognitive impairments, this population was expected to have a higher BR_{Fail} than NN or controls. Lastly, we anticipated no BR_{Fail} difference between NN and controls. It is worth noting that these two groups were not matched for demographic variables. As with all undergraduate samples, we expected them to represent a unique set of demographic variables (young, educated, high functioning) difficult to generalize to the general population. However, because EVIs are believed to be insensitive to the majority of demographic variables, and it is assumed that both groups are performing to their true ability, no significant differences were anticipated. In summary, we hypothesized the following gradient of BR_{Fail}:

$$EXP_{MAL} = HAD > MND = ANI > NN = control$$

In order to develop a more thorough understanding of how invalid performance might impact scores on cognitive tests within HIV+ individuals, members of the MND, ANI, and NN groups were merged and then divided according to the total number of EVIs failed. The relationship between the number of EVIs failed and the examinee's neuropsychological test

scores were then investigated. We hypothesized a strong negative linear relationship between number of EVI failures and performance on neuropsychological tests.

For the final part of our study, we examined the relationship between emotional functioning and performance validity. The MND, ANI, NN, and control group were merged and participants were regrouped according to their overall Patient Health Questionnaire (PHQ-9) scores. PHQ-9 scores were then compared to the number of EVIs failed to determine whether they may be a predictor of performance validity. It was hypothesized that those with higher PHQ-9 scores would have a greater number of EVI failures.

CHAPTER II

Methods

Participants

Sample 1: HIV+. The first sample was archival, using the data collected by Gomez, Power, Gill, and Fujiwara (2017). Their original research question asked whether risk-based decision making in an HIV+ sample correlated with a number of variables including neurocognitive performance. In the following section, a description of their cohort and methodology is provided.

Recruitment. All participants were diagnosed with HIV and recruited from the Southern Alberta Clinic (SAC) in Calgary, Alberta, Canada. Testing was completed between May 2013 and January 2016. All participants had normal or corrected-to-normal vision and hearing. Further, they all had sufficient English fluency and were capable of providing informed consent.

Demographic variables. Gender, sexual orientation, current/nadir CD4 and T-cell count, current/peak plasma HIV viral load, psychiatric and medical comorbidities, duration of HIV infection, cART status, cART side effects and polypharmacy, and cART regimen central nervous system (CNS)-penetration effectiveness rank were collected via chart review. The participant's age, ethnicity, country of origin, years of education, hepatitis C coinfection, and past/present substance use were collected via chart review and followed-up during the interview. Finally, cART adherence within the previous 5 days as well as perceived health and daily functioning were assessed during the interview.

HAND diagnosis. The "Frascati criteria" were used to determine HAND status. Diagnoses were further verified using the participant's medical record. HAND status could not be obtained if the participant had a history of head trauma with loss of consciousness exceeding

5 minutes, severe psychiatric or neurological disorders, or opportunistic CNS infection. At the time of testing, the participants spoke English fluently, and had greater than 9 years of education.

Number of participants. A total of 291 participants were enrolled in the study, 64 with neurocognitive impairments classified as HAND (25 ANI, 31 MNI, 8 HAD) and 227 NN.

Ethics approval. Data set collection was approved by the University of Calgary Conjoint Health Research Ethics Board (ethics ID: REB13-0615_REN2). All participants consented for their data to be used for research purposes. Ethical approval was also received from the University of Windsor Research Ethics Board.

Sample 2: Undergraduates. The second group of participants were prospectively collected undergraduates. The student sample provided control and EXP_{MAL} groups.

Recruitment. Undergraduate students were recruited from the University of Windsor Psychology Participant Pool. One screening question was asked prior to viewing the recruitment posting: "Are you 18 years of age or older?". If the answer was YES, the student was able to view the recruitment posting. The recruitment posting asked that participants identify as HIV- prior to signing up for the study. An email reminder was sent to students signed up for the study 48 hours prior to testing and a 24-hour cancellation notice was set. In order to optimize testing conditions, participants were asked to bring their glasses/contacts and/or hearing aid to the appointment. The battery took approximately 120 minutes to complete and participants received 2.5 credits for their involvement.

Inclusion criteria. All participants were 18 years of age or older and reported that they were HIV-.

Number of participants. A total of 74 undergraduate participants were recruited for the study. Two students were assigned to the control condition for each student assigned to EXP_{MAL}

(2 controls: 1 EXP_{MAL}). Only controls were included within the majority of our hypotheses, thus we wanted to maximize the sample size of this condition. Further, the 2:1 ratio was identified by G*Power as sufficient to observe group differences if they exist (Faul, Erdfelder, Lang, & Buchner, 2007). In total, fifty-one participants were randomly assigned to the control condition, while the remaining 23 were assigned to EXP_{MAL}.

Measures

Gomez, Gill, Power, and Fujiwara (2017). All testing was completed in a quiet, distraction-free room located in the SAC. After giving informed consent, neuropsychological testing followed a brief interview. All testing was completed by a trained researcher at the clinic.

HAND battery. The format of the HIV+ neurocognitive assessment was as follows: brief interview, D-KEFS verbal fluency (FAS/animals/boys names), Wide Range Achievement Test 4 (WRAT-4) Reading, Game of Dice, HVLT-R (learning trials 1-3, recognition trial), SDMT, WCST (64 cards), D-KEFS Trails 2 and 4, GPB (dominant hand, non-dominant hand), PHQ-9, HVLT-R (delayed recall).

Prospective undergraduate sample. The measures and procedures selected to be used in the undergraduate sample aimed to replicate the Gomez, Gill, Power, and Fujiwara (2017) study as closely as possible. However, the Game of Dice was removed as it adds to administration time, is an experimental measure without statistical norms, and is not involved in our research questions. One free-standing PVTs (WCT) was added to the battery to improve the classification accuracy of undergraduate participant profiles as valid or invalid (Iverson, Franzen, & McCracken, 1994).

Test administration. Testing was completed in a quiet, distraction-free room in the University of Windsor. Informed consent was obtained from all participants prior to

psychometric testing. Participants completed a brief questionnaire in order to obtain basic demographic information. Neuropsychological testing followed. All participants were fully debriefed following testing. The test sequence was as follows: brief interview, WCT (counterbalanced; either at the beginning or end of battery), WRAT-4 Reading, D-KEFS verbal fluency (FAS/animals/boys names), HVLT-R (learning trials 1-3), SDMT, D-KEFS Trails 2 and 4, GPB (dominant hand, non-dominant hand), HVLT-R (delayed recall, recognition), PHQ-9, and WCT (counterbalanced; either at the beginning or end of battery).

Demographic variables. Basic demographic information was collected during the questionnaire (e.g. gender, age, handedness, and years of education). Psychiatric and neurological histories were also collected (Appendix A).

Experimental malingering & control conditions. Undergraduates were randomly assigned into one of two conditions: EXP_{MAL} or control. Following the brief interview, participants were given a sealed envelope containing instructions on how to perform for the remainder of the experiment. Envelopes were quasi-counterbalanced to ensure random group selection. The envelope was used to prevent the researcher conducting the experiment from knowing the participant's condition and potentially biasing results. The instructions to controls requested they put forth their best effort while completing all tests. In contrast, participants in the EXP_{MAL} condition received detailed instructions on how to feign cognitive deficits in a pattern similar to that following a moderate to severe TBI. The given scenario has been previously used within our lab and was modelled after scenarios developed by DenBoer & Hall (2007) and Suhr & Gunstad (2000; Appendix B). The recommendations for simulation studies provided by Rogers (2008) were adhered to. Following the end of testing, a manipulation check was completed (Appendix C).

WRAT-4. Similar to the original study, participants completed the WRAT-4 Reading subtest. The WRAT-4 Reading subtest (blue version) is a list of 55 words, ordered according to difficulty, which the participant must read aloud. Pronunciation was scored as correct or incorrect (Wilkinson & Robertson, 2006).

PHQ-9. The PHQ-9 is a brief questionnaire measuring self-reported depressive symptoms over the past two weeks (Kroenke, Spitzer, & Williams, 2001).

Performance validity tests. D-KEFS verbal fluency, D-KEFS Trails, WCST, GPB, and HVLT-R all contain EVIs. Therefore, a total of 8 EVIs included within the original HAND battery were analyzed and tested in undergraduates (Table 1). The EVIs span a variety of cognitive domains and include both verbal and non-verbal measures. Conservative and liberal cutoff scores were chosen for each test to optimize specificity or sensitivity, respectively. Conservative cutoff scores aimed to have ≥ .90 specificity whereas liberal cutoff scores had improved sensitivity at the expense of slightly reduced specificity (≥ .84) (Boone, 2013; Larrabee, 2003).

Table 1
Neurocognitive Testing Battery

Name	Abbreviation	EVI	Reference
Letter Fluency	FAS	Yes	Delis, Kaplan, & Kramer, 2001
Category Fluency	Animals	Yes	Delis, Kaplan, & Kramer, 2001
WRAT-4 Reading	WRAT-4 Reading	No	Wilkinson & Robertson, 2006
HVLT-R	HVLT-R	Yes	Brandt & Benedict, 2001
Symbol Digit	SDMT	No	Smith. 1973
WCST 64 Card	WCST-64	Yes	Kongs, Thompson, Iverson, &
D-KEFS Trails	T2 & T4	Yes	Delis, Kaplan, & Kramer, 2001
Grooved Pegboard	GPB	Yes	Trites, 1977
PHQ-9	PHQ-9	No	Kroenke, Spitzer, & Williams, 2001

D-KEFS letter fluency. Participants were instructed to generate as many words as they could think of beginning with a specific letter (F, A, and S) in 60 seconds following some basic rules (cannot use proper names, numbers or the same word with different suffix; Delis, Kaplan,

& Kramer, 2001). Many of the EVIs within FAS are typically based on demographically adjusted T-scores (Curtis, Thompson, Greve, & Bianchini, 2008; Sugarman & Axelrod, 2015), which differ from age corrected scaled scores (ACSS) in the D-KEFS norms. Thus, raw scores were converted to T-scores using demographically adjusted norms by Heaton, Miller, Taylor, & Grant (2004). DelisSugarman and Axelrod (2015) found that an FAS T-score of ≤ 31 produced .90 specificity and .30 sensitivity, while an animal cutoff of T ≤ 33 had a .91 specificity and .42 sensitivity. Similarly, Curtis, Thompson, Greve, & Biachini (2008) suggested cutoff scores of ≤ 31 (.95 specificity, .27 sensitivity) and ≤33 (.90 specificity, .36 sensitivity) on FAS.

D-KEFS category fluency. Following the same structure as letter fluency, category fluency required the participant to list as many animals and boys names as they can within 60 seconds (Delis et al., 2001). Sugarman and Axelrod (2015) suggested a conservative raw score cutoff of ≤12 (sensitivity: .50; specificity: .90) and a liberal raw score cutoff of ≤ 13 (sensitivity: .55; specificity: .84) on animal fluency. The combined total of both categories was converted to an ACSS and used as a measure of cognitive ability.

HVLT-R. This is a test of auditory verbal learning and memory. Participants listened to a list of 12 words and were asked to recall as many as they could after each trial. For the purposes of our testing, three acquisition trials and a Yes/No recognition trial were administered (Brandt & Benedict, 2001). One study has published EVI cutoffs for the discrimination trial. They found correct responding during the discrimination trial ≤ 5 has a sensitivity of .93 but a specificity of .53, while the number of correct responses being ≤ 6 has a sensitivity of .74 and specificity of .84 (Sawyer, Testa, & Dux, 2017).

WCST. This test is a measure of concept formation and cognitive flexibility. Participants were asked to match each card, handed one at a time, to one of four key cards. Each card

contains 3 salient features (colour, form, number; Kongs, Thompson, Iverson, & Heaton, 2000). In the full 128 card version, FMS and the number of categories completed were included into a logistic regression equation that successfully differentiated credible from non-credible performance (Suhr & Boyer, 1999). FMS on the WCST is relatively insensitive to TBI and executive deficits in both adults (Heaton, Chelune, Talley, Kay, & Curtis, 1993; Jodzio & Biechowska, 2010) and children (Lichtenstein, Erdodi, Rai, Mazur-Mosiewicz, & Flaro, 2016). In contrast to healthy undergraduates and those with credible TBI, experimental malingerers and patients with TBI seeking compensation had more than double the number of FMS errors (≥ 2; King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Suhr & Boyer, 1999). Larrabee (2003) reported .87 specificity and .48 sensitivity for FMS errors ≥ 2. Across a variety of non-litigating clinical populations, the mean FMS was consistently <1. However, FMS as an EVI has been calibrated using the 128-card version but the current study used the 64-card version. As such, the theoretical probability of FMS errors is lower (half). Therefore, at conventional cutoffs, the FMS in the 64-card version is expected to have higher specificity.

D-KEFS Trails 2 and 4. When completing Trail 2, examinees were instructed to connect circles containing numbers in increasing order using a pencil. Trail 4 requires letter-number sequencing: when connecting letters to numbers, examinees are asked to alternate between numbers in increasing and letters in alphabetical order. Trail 2 measures processing speed and simple visual attention. Trail 4 measures cognitive flexibility and divided attention (Delis et al., 2001). Although the original Trail Making Test has been validated as a PVT (Busse & Whiteside, 2012; Iverson, Lange, Green, & Franzen, 2002; Ruffolo, Guilmette, & Willis, 2000; Shura, Miskey, Rowland, Yoash-Gantz, & Denning, 2016), there has been no research exploring the utility of D-KEFS TMT in detecting non-credible performance. Our research group recently

investigated the potential of D-KEFS Trails to function as EVIs (Erdodi, Hurtubise, et al., 2018). We found that a Trails 2 ACSS cutoff of ≤5 met minimum specificity standards (.85 - .88) with a sensitivity of .43-.57. At the more conservative cutoff of ≤3, specificity improved (.87-.93) at the expense of sensitivity (.26-.38). On Trails 4, cutoffs of ≤4 (specificity: .88, sensitivity: .45-.57) and ≤ 1 (specificity: .90-.93, sensitivity: .27-.48) produced good combinations of specificity and sensitivity.

Grooved Pegboard. This test is a measure of fine motor speed and requires participants to rotate pegs into peg holes using their dominant and, later, their non-dominant hand. Speed and the number of pegs dropped were recorded (Trites, 1977). A dominant hand T-score cutoff ≤ 29 produced a sensitivity of .61 and a specificity of .90 when groups were classified according to performance on the words subtest of the Warrington's Recognition Memory Test and two other composite measures of performance validity. In contrast, a dominant hand T-score cutoff of ≤ 25 had a sensitivity of .52 and a specificity of .96. The same cutoffs produced good combinations of sensitivity and specificity for the non-dominant hand (liberal: .65 and.89; conservative: .50 and .96; Erdodi, Seke, et al., 2017).

Statistical Analysis

All statistical procedures were performed using Statistical Package for the Social Sciences (SPSS) version 22.0. Descriptive statistics (mean, SD, range, skew, kurtosis) for demographic variables were reported in all four groups. These variables were compared using a *t*-test for continuous variables (e.g. age, education) and a chi-square test of independence for categorical variables (e.g. gender). For all hypotheses, BR_{Fail} across groups was compared at liberal and conservative cutoffs separately.

Hypothesis 1: EVI failure will vary according to group membership (EXP_{MAL}= HAD > MND = ANI > NN = CON). It was expected that this prediction would remain true regardless of whether the dependent variable was the total number of EVIs failed (continuous) or BR_{Fail} (categorical: *Pass/Fail*). For each EVI, participants were scored as either passing or failing according to the predetermined cutoff. Each participant received a total score of the number of EVIs they failed. As this was a continuous variable with a maximum score of 8, a between-subjects analysis of variance (ANOVA) was conducted and a comparison was made between the six groups. Assumptions checked included normality, equal variance, and independence of groups. Post-hoc contrasts were uncorrected post-hoc tests. An effect size estimate (Cohen's *d*) was computed for significant contrasts.

BR_{Fail} was compared between the six groups by looking at whether groups varied in the number of individuals failing ≤1, 2, and ≥3 EVIs at the liberal cutoffs and ≤2, 3, and ≥4 EVIs at the conservative cutoffs. These analyses were done to compare proportions between groups. To allow for easier clinical interpretation, risk ratios were computed, followed by χ^2 analysis to determine statistical significance.

Hypothesis 2: Those failing a greater number of EVIs will perform more poorly on neuropsychological tests. To test this hypothesis, the MND, ANI and, NN groups were merged together and this large pool was divided according to the number of EVIs each participant failed. All neurocognitive measures that doubled as EVIs were removed from this analysis to reduce the effects of collinearity. For the purpose of our analysis, the independent variable had 3 levels at the liberal cutoffs: ≤2 EVIs failed, 3 EVI failed, and ≥4 EVIs failed and 3 levels at the conservative cutoffs: ≤1 EVIs failed, 2 EVI failed, and ≥3 EVIs failed. ANOVAs were

conducted to compare the performance across these three groups on six measures of cognitive ability.

Hypothesis 3: Regardless of HIV status, individuals who have elevated PHQ-9 scores will fail more EVIs than those who do not. Using their PHQ-9 scores, all participants except the EXP_{MAL} and HAD groups were divided into none (0), minimal (1-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (≥ 19) depression. In a manner similar to Hypothesis 1, BR_{Fail} was conducted for ≤2, 3, and ≥4 EVIs at the liberal cutoffs and ≤1, 2, and ≥3 EVIs at the conservative cutoffs using a χ^2 analysis. Additionally, the total number of failed EVIs between participants was calculated and compared between the six groups (none, minimal, mild, moderate, moderately severe, and severe). As there were six groups to compare, an ANOVA was conducted. An additional ANOVA was conducted assessing the relationship between EVI failure and depressive symptom severity as indicated by question 10 on the PHQ-9 (not very difficult, somewhat difficult, very difficult, and extremely difficult).

CHAPTER III

Results

Data Cleaning

Prior to data analysis, all relevant variables were scanned using descriptive statistics to identify human error in data scoring or entry. All data were deemed acceptable. Within the HIV+ dataset, no cases were removed. A total of 9 cases were removed from the student dataset and subsequent analysis. Of these, 2 were removed due to inadequate demographic information that prevented comparison of their performance to age-matched controls. One participant in the EXP_{MAL} group reported "not pretending" to have a head injury or following instructions. Additionally, 6 participants were excluded from subsequent analysis because they did not complete all 8 EVIs. The final sample size was 346 participants.

Testing of Assumptions

Skewness and Kurtosis. T-tests and ANOVAs assume that data are normally distributed. In order to test for the normal distribution of data, skewness and kurtosis values were assessed for all continuous variables (e.g. FAS T-Score, Animals T-Score, Verbal Fluency Boys Names, HVLT-R Total Learning, HVLT-R Delayed Recall, WCST Categories Completed, D-KEFS Trails 2 and 4, GPB Dominant and Non-Dominant Hand, and PHQ-9 Total Score). None of the variables produced skewness or kurtosis values outside the acceptable range of +2 and -2 (Pituch & Stevens, 2016).

Equality of Variance. Equal variance across populations was tested using the Levene's Test of Equality of Variances for each variable and is reported below.

Independence of Observations. The only assumption of χ^2 test of independence is the independence of observations. All groups contain greater than 5 cases and participants were

placed into a single group. Further, it is not believed that the group membership of one participant influenced the group membership of another.

Main Analyses

Demographic Variables. T-tests were conducted to determine whether demographic differences exist between HIV+ and student participants. Overall, the undergraduate sample was younger ($M_{student}$ = 22.1, $SD_{student}$ = 5.0) than the HIV+ sample (M_{HIV+} = 47.4, SD_{HIV+} = 10.9; t(354) = 28.45, $p < .001$, d = 2.98). The student sample had completed more years of education ($M_{student}$ = 14.6, $SD_{student}$ = 1.1) than those in the HIV+ dataset (M_{HIV+} = 14.1, SD_{HIV+} = 2.5; t(354) = -2.42, p = .016, d = 0.26).

Sample differences in handedness and gender were assessed using χ^2 tests of independence. Handedness did not differ between the two samples (χ^2 = 3.51, p = .173). The proportion of female and male participants varied between samples (χ^2 = 158.6, $p < .001$), with a greater proportion of female participants in the student (86.2%) than HIV+ (11.3%) sample.

Validating EVIs in the HAND Battery using a student sample. A total of eight established EVIs were included within the HAND battery. Predetermined cutoffs, outlined in the methods section, did not provide adequate sensitivity and specificity for the undergraduate sample. Therefore, an exploratory analysis was conducted. Beginning with the proposed cutoffs, the sensitivity and specificity of alternative cutoffs were also computed. The liberal and conservative cutoffs used for hypothesis testing were those producing specificity nearest to .84 and .90 respectively (Table 2). Two criterion measures were used to determine sensitivity and specificity: (1) EXP_{MAL} vs. Controls and (2) WCT raw scores > 47 (pass) vs. ≤ 47 (*Fail*; Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009).

Table 2

Liberal and Conservative Cutoffs for each EVI as determined by sensitivity and specificity of each EVI against criterion PVT in the sample

EVI	Scale	Liberal	Conservative
HVLT-R RD	Raw score	≥ 8	≥ 7
WCST FMS	Raw score	≤ 2	≤ 1
D-KEFS Trail 2	ACSS	≥ 6	≥ 5
D-KEFS Trail 4	ACSS	≥ 7	≥ 6
GPB DOM	T-score	≥ 31	≥ 29
GPB ND	T-score	≥ 33	≥ 31
FAS	T-score	≥ 33	≥ 31
Animals	T-score	≥ 33	≥ 29

Note. EVI: Embedded Validity Indicator; HVLT-R RD: Hopkins Verbal Learning Test – revised recognition discrimination (Sawyer, Testa, & Dux, 2017); WCST FMS: Wisconsin Card Sorting Test failure to maintain set (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999). D-KEFS T2 & T4: Delis-Kaplan Executive Functioning System, Trails 2 and 4 (Erdodi, Hurtubise, et al., 2018); GPB DOM: Grooved Pegboard Test dominant hand (Erdodi, Seke, et al., 2017); GPB ND: Grooved Pegboard Test non-dominant hand (Erdodi, Seke, et al., 2017); FAS & Animals: Delis-Kaplan executive functioning system (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Area under the curve (AUC) provides an objective measure of overall classification accuracy (i.e. determining whether a profile is valid or invalid). AUC may be classified as acceptable (.70-.79), excellent (.80-.89), or outstanding (≥ .90; Hosmer & Lemeshow, 2000). With the exception of WCST FMS, all EVIs fell within or above the acceptable range (Table 3). WCST FMS AUC values fell well below the acceptable level (.64) but within the acceptable range (.75) when using the EXP_{MAL} criterion and WCT as the PVT criterion, respectively. D-KEFS T2, FAS, and Animals produced acceptable classification accuracy. HVLT-R RD and GPB Dom had excellent classification accuracy. EVI classification varied depending on the criterion measure used for D-KEFS T4 and GPB ND. D-KEFS T4 AUC acceptable classification accuracy using the WCT and excellent classification accuracy when EXP_{MAL} was the criterion PVT. In contrast, excellent classification accuracy was observed with GPB ND using EXP_{MAL} criterion but outstanding accuracy with WCT as the criterion.

Table 3
Area *Under the Curve and Confidence Intervals of Select EVI Validity Cutoffs against Various Criterion PVTs*

	Criterion PVT			
	EXP_{MAL}		WCT	
	AUC	95% CI	AUC	95% CI
HVLT-R RD	.81	.69-.93	.88	.76-1.00
WCST FMS	.64	.48-.80	.75	.58-.91
D-KEFS T2	.79	.65-.92	.73	.58-.88
D-KEFS T4	.83	.72-.95	.75	.61-.89
GPB Dom	.82	.70-.94	.84	.73-.96
GPB ND	.80	.67-.92	.92	.83-1.00
FAS	.71	.57-.85	.70	.56-.84
Animals	.76	.63-.89	.76	.60-.92

Note. EVI: Embedded Validity Indicator; AUC: Area under the curve; 95% CI: 95% confidence interval; WCT: Word Choice Test (Pass: WCT score > 47; Fail: WCT score ≤47 (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009); HVLT-R RD: Hopkins Verbal Learning Test – Revised recognition discrimination (Sawyer, Testa, & Dux, 2017); WCST FMS: Wisconsin Card Sorting Test failure to maintain set (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999). D-KEFS T2 & T4: Delis-Kaplan Executive Functioning System, Trails 2 and 4 (Erdodi, Hurtubise, et al., 2018); GPB DOM: Grooved Pegboard Test dominant hand (Erdodi, Seke, et al., 2017); GPB ND: Grooved pegboard non-dominant hand (Erdodi, Seke, et al., 2017); FAS: Letter fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals: Category fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Table 4 presents the sensitivity and specificity for the liberal and conservative cutoffs of EVIs within the HAND battery. The classification accuracy for published cutoffs on seven of the eight EVIs within the HAND battery hovered around the *Larrabee limit:* .50 sensitivity at .90 specificity (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014) against EXP_{MAL} and WCT as criterion measures. Specificity values were generally lower against the WCT than EXP_{MAL}. This pattern of findings is likely an artifact of differences in the BR_{Fail} (21.9% vs. 32.3%). Overall, classification accuracy was similar across cutoffs between the two criterion measures.

Table 4

BR_{Fail} *, Sensitivity and Specificity of Select EVI Validity Cutoffs against Various Criterion PVTs*

			Criterion PVTs			
			EXP_{MAL}		WCT	
			SENS	SPEC	SENS	SPEC
EVI	Cutoff	BR_{Fail}	32.3		21.9	
HVLT-R RD	≤8	17.2	.42	.96	.64	.96
	≤7	14.1	.38	.98	.57	.98
WCST FMS	≥1	26.6	.43	.82	.64	.84
	≥2	9.4	.24	.98	.29	.96
D-KEFS T2	≤6	26.6	.57	.89	.43	.78
	≤5	14.1	.29	.93	.28	.90
D-KEFS T4	≥7	32.8	.67	.84	.50	.72
	≥6	26.6	.67	.93	.50	.80
GPB Dom	≤31	31.3	.67	.86	.64	.78
	≤29	23.4	.57	.93	.64	.88
GPB ND	≤33	29.7	.62	.86	.93	.88
	≤31	23.4	.52	.91	.78	.92
FAS	≤33	23.4	.43	.86	.36	.80
	≤31	18.8	.43	.93	.36	.86
Animals	≤33	20.3	.43	.91	.57	.91
	≤29	14.1	.33	.95	.43	.94

Note. EVI: Embedded validity indicator; BR_{Fail} : Base rate of failure (% of the sample that failed a given cutoff); WCT: Word Choice Test (Pass: WCT score > 47; Fail: WCT score ≤47 (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009); HVLT-R RD: Hopkins Verbal Learning Test – Revised recognition discrimination raw score (Sawyer, Testa, & Dux, 2017); WCST FMS: Wisconsin Card Sorting Test failure to maintain set raw score (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999). D-KEFS T2 & T4: Delis-Kaplan Executive Functioning System, Trails 2 and 4 age-corrected scaled score (Erdodi, Hurtubise, et al., 2018); GPB DOM: Grooved Pegboard Test dominant hand demographically adjusted T-score (Erdodi, Seke, et al., 2017); GPB ND: Grooved Pegboard Test non-dominant hand demographically adjusted T-score (Erdodi, Seke, et al., 2017); FAS: Letter Fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals: Category Fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015)

To further quantify the utility of each EVI, t-tests were conducted to compare performance for each criterion PVT (Table 5). Significant differences were observed between controls and EXP_{MAL} as well as between Pass and Fail of the WCT. Standard interpretation of Cohen's d suggests scores greater than .2 as small, greater than .5 as moderate, and greater than .8 as large effect sizes. Only WCST FMS had an effect size in the small range (d = .28). The remaining tests had moderate to large effects.

Table 5

The Effect of Invalid Performance on Various EVIs in the Student Sample

	Criterion PVTs													
	Experimental Malingering							WCT						
EVI	IV	*M*	*SD*	*t*	*p*	*d*		WCT	*M*	*SD*	*t*	*p*	*d*	
HVLT-R RD	NC	10.8	*1.3*	3.80	*.001*	1.12	†	Pass	10.8	*1.3*	4.26	*.001*	1.60	†
	EM	7.1	*4.5*					Fail	5.4	*4.6*				
WCST FMS	NC	.2	*.5*	-2.34	*.028*	.28	†	Pass	.2	*.7*	-2.7	*.014*	.98	†
	EM	.9	*1.2*					Fail	1.1	*1.1*				
D-KEFS T2	NC	9.8	*2.7*	4.47	*<.001*	1.10		Pass	9.2	*3.2*	2.89	*.005*	.84	
	EM	6.1	*3.9*					Fail	6.2	*3.9*				
D-KEFS T4	NC	9.5	*2.2*	4.88	*<.001*	1.40	†	Pass	8.8	*2.9*	2.79	*.012*	.87	†
	EM	5.5	*3.4*					Fail	5.9	*3.7*				
GPB Dom	NC	44.5	*11.2*	4.78	*<.001*	1.24		Pass	43.3	*12.2*	4.50	*<.001*	1.41	
	EM	29.9	*12.3*					Fail	27.2	*10.5*				
GPB ND	NC	43.5	*9.0*	4.26	*<.001*	1.08		Pass	43.4	*9.3*	5.88	*<.001*	1.86	
	EM	32.3	*11.6*					Fail	27.2	*8.1*				
FAS	NC	42.1	*9.9*	2.74	*.008*	.72		Pass	41.3	*11.0*	2.19	*.032*	.74	
	EM	34.8	*10.3*					Fail	34.4	*7.4*				
Animals	NC	44.5	*11.2*	3.79	*<.001*	1.24		Pass	44.0	*9.4*	3.69	*<.001*	1.03	
	EM	29.9	*12.3*					Fail	32.9	*12.0*				

Note. EVI: Embedded validity indicator; NC: Normal controls; EM: Experimental malingerer; WCT: Word Choice Test (Pass: WCT score > 47; Fail: WCT score ≤47 (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009)); HVLT-R RD: Hopkins Verbal Learning Test – Revised recognition discrimination (Sawyer, Testa, & Dux, 2017); WCST FMS: Wisconsin Card Sorting Test failure to maintain set (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999). D-KEFS T2 & T4: Delis-Kaplan Executive Functioning System, Trails 2 and 4 (Erdodi, Hurtubise, et al., 2018); GPB DOM: Grooved Pegboard Test dominant hand (Erdodi, Seke, et al., 2017); GPB ND: Grooved Pegboard Test non-dominant hand (Erdodi,Seke et al., 2017); FAS: Letter fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals: Category fluency T-score (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

†: Levene's test of homogeneity of variance significant at p-value <.05

Although it is useful to understand each EVI individually, cumulative EVI failure\e is more clinically relevant. For this reason, AUC and 95% confidence intervals were also calculated by summing the total number of EVIs failed at liberal and conservative cutoffs (Table 6). Regardless of criterion PVT, liberal cutoffs produced excellent classification accuracy (.88-.89). Although WCT AUC suggested acceptable classification accuracy using conservative cutoffs (.78), the EXP_{MAL} criterion suggested that it is outstanding (.92).

Table 6

AUC and 95% CI of Select Levels of Failure against Various Criterion PVTs

	Criterion PVT			
	EXP_{MAL}		WCT	
Number of EVI Failures	AUC	95% CI	AUC	95% CI
Liberal	.89	.79-.98	.88	.80-.98
Conservative	.92	.84-.99	.78	.76-.99

Note: EXP_{MAL}: Experimental malingering condition; WCT: Word Choice Test (Pass: WCT score > 47; Fail: WCT score ≤47 (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009)); AUC: Area under the curve; Liberal: EVIs failed at the liberal cutoff (Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8; Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2; Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6; D-KEFS Trail 4 ≤ 7; Grooved Pegboard Test (GPB) dominant hand ≤ 31; GPB non-dominant hand ≤ 33; FAS ≤ 33; Animals ≤ 33); Conservative: EVIs failed at the conservative cutoff (HVLT-R RD ≤ 7; WCST FMS ≥ 1; D-KEFS Trail 2 ≤ 5; D-KEFS Trail 4 ≤ 6; GPB dominant hand ≤ 29; GPB non-dominant hand ≤ 31; FAS ≤ 31; Animals ≤ 29).

BR_{Fail}, specificity, and sensitivity were calculated using cumulative failure cutoffs for liberal and conservative EVIs (Table 7). As the number of EVIs failed increased, cutoffs produced greater specificity at the expense of sensitivity. At the liberal and conservative cutoffs of ≥4 and ≥3, respectively, sensitivity and specificity approached the Larrabee limit (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014). Again, specificity values were greater when using EXP_{MAL} as the criterion PVT rather than WCT.

Table 7
Sensitivity and Specificity of the Cumulative EVI Failures at Select Cutoffs against Various Criterion PVTs

			Criterion PVT			
			EXP_{MAL}		WCT	
		BR_{Fail}	32.3		21.9	
Level of Cutoff	Number Failures		SENS	SPEC	SENS	SPEC
Liberal	≥1	67.7	.95	.46	1.00	.40
	≥2	47.7	.86	.71	.93	.64
	≥3	30.8	.81	.76	.79	.82
	≥4	23.1	.57	.93	.64	.88
	≥5	13.8	.43	1.00	.50	.96
Conservative	≥1	47.7	.95	.75	.93	.64
	≥2	32.3	.91	.81	.80	.79
	≥3	23.1	.57	.93	.71	.90
	≥4	18.5	.48	.96	.57	.92
	≥5	9.2	.29	1.00	.36	.98

Note: Embedded validity indicator; BR_{Fail} : Base rate of failure (% of the sample that failed a given cutoff); EXP_{MAL}: Experimental malingering condition; WCT: Word Choice Test (Pass: WCT score > 47; Fail: WCT score ≤47 (Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Pearson, 2009); Liberal cutoffs (Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8; Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2; Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6; D-KEFS Trail 4 ≤ 7; Grooved Pegboard Test (GPB) dominant hand ≤ 31; GPB non-dominant hand ≤ 33; FAS ≤ 33; Animals ≤ 33); Conservative cutoffs (HVLT-R RD ≤ 7; WCST FMS ≥ 1; D-KEFS Trail 2 ≤ 5; D-KEFS Trail 4 ≤ 6; GPB dominant hand ≤ 29; GPB non-dominant hand ≤ 31; FAS ≤ 31; Animals ≤ 29).

EVI Failure and Group Membership. A between-subjects ANOVA was conducted to determine whether the number of liberal and conservative EVIs failed differed between the six groups (Table 8 and 9). The pattern of EVI failure was consistent between liberal and conservative cutoffs. For both, post-hoc analysis identified that EXP_{MAL} and HAD groups failed significantly more EVIs than the other 4 groups but did not differ from each other. Those with MND failed more EVIs than controls, NN, or ANI. ANI failed significantly more EVIs than controls or NN. The number of EVIs failed did not differ between controls and NN.

Table 8
Comparison of the total number of Liberal EVIs failed across groups

	Total EVIs failed						
	n	*M*	SD	*F*	*df*	*p*	η_p^2
Control	44	1.0	1.2	50.58	5	.000	.42
EXP_{MAL}	21	4.2	2.4				
NN	227	0.82	0.9				
ANI	25	1.6	1.4				
MND	31	2.3	1.6				
HAD	8	4.4	1.1				

Note. EVI: Embedded validity indicator; EXP_{MAL}: Experimental malingerers; NN: HIV+ and neurocognitively normal; ANI: Asymptomatic neurocognitively impaired; MND: Mild neurocognitive disorder; HAD: HIV-associated dementia.

Table 9
Comparison of the total number of Conservative EVIs failed across groups

	Total EVIs failed						
	n	*M*	SD	*F*	*df*	*p*	η_p^2
Control	44	0.5	1.0	60.33	5	.000	.46
EXP_{MAL}	21	3.4	2.2				
NN	227	0.4	0.6				
ANI	25	1.0	0.9				
MND	31	1.7	1.5				
HAD	8	3.5	1.4				

Note. EVI: Embedded validity indicator; EXP_{MAL}: Experimental malingerers; NN: HIV+ and neurocognitively normal; ANI: Asymptomatic neurocognitively impaired; MND: Mild neurocognitive disorder; HAD: HIV-associated dementia.

Table 10 presents the frequency distribution of the different profile validity classifications across groups. Regardless of cutoff, the majority of controls, NN, ANI, and MND participants produced valid profiles. In contrast, the majority of EXP_{MAL} and HAD individuals produced invalid profiles.

Table 10
Profile Validity Distribution of All Participants by Group

		Liberal			Conservative		
Condition	n	*Valid* (≤2)	*Borderline* (3)	*Invalid* (≥4)	*Valid* (≤1)	*Borderline* (2)	*Invalid* (≥3)
Control	44	40	1	3	40	1	3
EXP_{MAL}	21	4	5	12	5	4	12
NN	227	219	5	3	213	12	2
ANI	25	20	2	3	18	6	1
MND	31	19	6	6	18	4	9
HAD	8	0	2	6	1	1	6
Total	356	302	21	33	295	28	33

Note. EXP_{MAL}: Experimental malingerers; NN: HIV+ and neurocognitively normal; ANI: Asymptomatic neurocognitively impaired; MND: Mild neurocognitive disorder; HAD: HIV-associated dementia; Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Conservative cutoffs: HVLT-R RD ≤ 7 (Sawyer, Testa, & Dux, 2017); WCST FMS ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); D-KEFS Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); GPB dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Significant differences were observed between group membership and the proportion of individuals failing EVIs at all liberal and conservative cutoffs investigated (Table 11 and 12). Individuals in the EXP_{MAL} and HAD groups were consistently more likely to fail than pass cumulative EVI cutoffs, as indicated by risk ratios. In contrast, NN and controls were more likely to pass than fail EVIs at all cutoffs.

Table 11

Percentage of Failure, Risk Ratio, and χ^2 analysis of all groups at ≥2, ≥3, and ≥4 liberal EVIs failed

Criteria	Condition	%	RR	χ^2	*p*	ϕ^2
≥2	Control	29.5	0.4	89.71	< *.001*	.25
	EXP_{MAL}	85.7	6.0			
	NN	17.6	0.2			
	ANI	52.0	1.1			
	MND	67.7	2.1			
	HAD	100.0	-			
≥3	Control	9.1	0.1	146.72	< *.001*	.41
	EXP_{MAL}	76.2	3.2			
	NN	3.5	<.1			
	ANI	20.0	.3			
	MND	38.7	.6			
	HAD	100.0	-			
≥4	Control	6.8	.1	119.66	<*.001*	.34
	EXP_{MAL}	57.1	1.3			
	NN	1.3	<.1			
	ANI	12.0	.1			
	MND	19.4	.2			
	HAD	75.0	3.0			

Note. EXP_{MAL}: Experimental malingerers; NN: HIV+ and neurocognitively normal; ANI: Asymptomatic neurocognitively impaired; MND: Mild neurocognitive disorder; HAD: HIV-associated dementia; Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Table 12

Percentage of Failure, Risk Ratio, and χ^2 analysis of all groups at ≥1, ≥2, and ≥3 Conservative EVIs failed

Criteria	Condition	%	RR	χ^2	*p*	ϕ^2
≥1	Control	25.0	.3	79.58	<.*001*	.22
	EXP_{MAL}	95.2	20			
	NN	29.1	.4			
	ANI	68.0	2.1			
	MND	77.4	3.4			
	HAD	100.0	-			
≥2	Control	9.1	0.1	123.3	<.*001*	.35
	EXP_{MAL}	81.0	4.3			
	NN	6.2	.1			
	ANI	28.0	.4			
	MND	41.9	.7			
	HAD	87.5	7			
≥3	Control	6.8	.1	132.9	<.*001*	.37
	EXP_{MAL}	57.1	1.3			
	NN	0.9	<.1			
	ANI	4.0	<.1			
	MND	29.0	.4			
	HAD	75.0	3			

Note. EXP_{MAL}: Experimental malingerers; NN: HIV+ and neurocognitively normal; ANI: Asymptomatic neurocognitively impaired; MND: Mild neurocognitive disorder; HAD: HIV-associated dementia; Conservative cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 7 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard test (GPB) dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

EVI Failure and Neuropsychological Test Performance. To test the influence of EVI failure on neuropsychological test performance in the HIV+ population, the undergraduate sample was removed. Further, the HAD group data was also removed as this population is known to have severe and genuine impairment that would increase EVI failure and reduce test performance (Davis, 2018). The remaining 4 groups (controls, NN, ANI, and MND) were merged and regrouped according to the total number of liberal and conservative EVIs failed. ANOVAs were conducted to explore group differences according to the total number of EVIs

failed (Table 13 and Table 14). Across all neuropsychological tests, the *valid* (liberal: ≤2; conservative: ≤1) and *invalid* (liberal: ≥4; conservative: ≥3) profile groups performed differently, with moderate to large effect sizes (*d* = .60-1.64). The *borderline* group (liberal: 3; conservative: 2) had significantly poorer performance than the *valid* group on verbal fluency categories, HVLT-R total learning, HVLT-R delayed recall (liberal cutoff only), WCST categories completed, and SDMT. Neuropsychological test performance between the *borderline* and *invalid* groups never significantly differed.

Table 13
Comparison of Neuropsychological Test Performance across Levels of Liberal EVI Failure

	Number of EVIs Failed										
	≤2		3		≥4						
	n = 258		n = 13		n = 12						
NP Test	*M*	SD	*M*	SD	*M*	SD	*F*	*p*	η_p^2	Sig. *post hocs**	*d*
WRAT-4	105.0	12.7	94.8	14.1	91.4	13.1	10.03	<.*001*	.07	≤2 vs. 3	.77
										≤2 vs. ≥4	1.01
VFlu Cat	10.9	3.3	6.9	3.4	6.3	2.7	20.14	<.*001*	.13	≤2 vs. 3	1.21
										≤2 vs. ≥4	1.56
HVLT TL	42.1	10.8	31.0	6.5	29.9	7.7	13.81	<.*001*	.10	≤2 vs. 3	1.24
										≤2 vs. ≥4	1.30
HVLT DR	41.7	12.6	29.8	12.7	30.1	11.1	9.97	<.*001*	.07	≤2 vs. 3	1.02
										≤2 vs. ≥4	1.07
WCST Cat	3.3	1.4	2.1	1.4	2.4	0.8	7.55	.*001*	.05	≤2 vs. 3	.91
										≤2 vs. ≥4	.81
SDMT	0.04	1.0	-1.4	1.2	-1.4	.8	22.94	<.*001*	.14	≤2 vs. 3	1.29
										≤2 vs. ≥4	1.57

Note. EVI: Embedded validity indicator; WRAT-4: Wide Range Achievement Test – reading subtest scaled score (Wilkinson & Robertson, 2006); VFlu Cat: Verbal fluency categories scaled score (Delis, Kaplan, & Kramer, 2001); HVLT DR: Hopkins Verbal Learning Test - Revised delayed recall T-score (Brandt & Benedict, 2001); HVLT TL: Hopkins Verbal Learning Test - Revised total learning T-Score (Brandt & Benedict, 2001); WCST Cat: Wisconsin Card Sorting Test categories completed (Kongs, Thompson, Iverson, & Heaton, 2000); SDMT: Symbol Digit Modalities Test z-score (Smith. 1973); Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

* Least significant difference (uncorrected *t*-tests)

Table 14

Comparison of Neuropsychological Test Performance across Levels of Conservative EVI Failure

	Number of EVIs Failed										
	≤1		2		≥3						
	n = 249		n = 22		n = 12						
	(88%)		(8%)		(4%)						
NP Test	*M*	SD	*M*	SD	*M*	SD	*F*	*p*	η_p^2	*Sig. post hocs**	*d*
WRAT-4	104.9	12.8	99.5	12.1	92.5	16.4	6.60	*.002*	.05	≤1 vs. ≥3	*.84*
VFlu Cat	11.0	3.3	8.1	3.4	5.9	2.4	20.28	*<.001*	.13	≤1 vs. 2	*.85*
										≤1 vs. ≥3	*1.77*
HVLT TL	42.2	10.8	36.1	9.8	27.8	6.9	13.18	*<.001*	.09	≤1 vs. 2	*.59*
										≤1 vs. ≥3	*1.59*
										2 vs. ≥3	*.98*
HVLT DR	41.6	12.8	36.2	12.3	28.8	10.5	7.37	*.001*	.05	≤1 vs. ≥3	*1.10*
WCST Cat	3.3	1.4	2.7	1.5	2.3	0.8	5.49	*.005*	.04	≤1 vs. 2	*.45*
										≤1 vs. ≥3	*.97*
SDMT	0.1	1.0	-0.9	1.2	-1.7	1.0	25.32	*<.001*	.15	≤1 vs. 2	*.86*
										≤1 vs. ≥3	*1.83*
										2 vs. ≥3	*.78*

Note. WRAT-4: Wide Range Achievement Test – reading subtest scaled score (Wilkinson & Robertson, 2006); VFlu Cat: Verbal Fluency categories scaled score (Delis, Kaplan, & Kramer, 2001); HVLT DR: Hopkins Verbal Learning Test revised delayed recall T-score (Brandt & Benedict, 2001); HVLT TL: Hopkins Verbal Learning Test revised total learning T-Score (Brandt & Benedict, 2001); WCST Cat: Wisconsin Card Sorting Test categories completed (Kongs, Thompson, Iverson, & Heaton, 2000); SDMT: Symbol Digit Modalities Test z-score (Smith. 1973); Conservative cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 7 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard test (GPB) dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

EVI Failure and Depression Symptom Endorsement. For these analyses, EXP_{MAL} and HAD groups were removed. The remaining four groups (controls, NN, ANI, MND) were merged and divided into six new groups according to their raw PHQ-9 scores. Between subjects ANOVAs were used to determine whether PHQ-9 symptom ratings influenced EVI failure at

liberal or conservative cutoffs (Table 15 and 16). PHQ-9 raw scores did not influence EVI failure ($p > .05$).

Table 15
Comparison of the total number of Liberal EVIs failed across PHQ-9 Score groups

	Total EVIs failed						
	n	*M*	SD	F	*df*	*p*	η_p^2
None	49	.92	1.1	1.12	5	*.351*	.017
Minimal	118	.92	1.0				
Mild	87	1.3	1.3				
Moderate	51	1.1	1.3				
Moderately Severe	13	1.2	1.1				
Severe	8	1.4	1.3				

Note. EVI: Embedded validity indicator; None: PHQ-9 score of 0; Minimal: PHQ-9 score of 1-4; Mild PHQ-9 score of 5-9; Moderate: PHQ-9 score of 10-14; Moderately Severe: PHQ-9 score of 15-19; Severe: PHQ-9 score greater than 19 (Kroenke, Spitzer, & Williams, 2001); Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Table 16
Comparison of the total number of Conservative EVIs failed across PHQ-9 Score groups

	Total EVIs failed						
	n	*M*	SD	F	*df*	*p*	η_p^2
None	49	0.5	0.8	1.415	5	*0.22*	0.22
Minimal	118	0.4	0.7				
Mild	87	0.7	1.1				
Moderate	51	0.7	1.1				
Moderately Severe	13	0.7	0.8				
Severe	8	0.8	1.0				

Note. EVI: Embedded validity indicator; None: PHQ-9 score of 0; Minimal: PHQ-9 score of 1-4; Mild PHQ-9 score of 5-9; Moderate: PHQ-9 score of 10-14; Moderately Severe: PHQ-9 score of 15-19; Severe: PHQ-9 score greater than 19 (Kroenke, Spitzer, & Williams, 2001); Conservative cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 7 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard test (GPB) dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

A Chi-square test of independence was calculated comparing the proportion of individuals with invalid performance for each of the 6 PHQ-9 raw score groups at six cumulative cutoffs (Table 17 and 18). No significant interactions were observed between PHQ-9 raw scores and the level of liberal or conservative EVI failure.

Table 17

Proportion Failing, Risk Ratio, and χ^2 analysis according to Depressive Symptoms reported at ≥2, ≥3, and ≥4 Liberal EVIs failed

# of EVI failures	Symptom endorsement	%	RR	χ^2	*p*	ϕ^2
≥2	None	22.4	.3	3.79	*.580*	.11
	Minimal	23.7	.3			
	Mild	33.3	.5			
	Moderate	23.5	.3			
	Mod. Sev	30.8	.4			
	Severe	37.5	.6			
≥3	None	6.1	.1	2.70	*.747*	.01
	Minimal	6.8	.1			
	Mild	11.5	.1			
	Moderate	9.8	.1			
	Mod. Sev	15.4	.2			
	Severe	12.5	.1			
≥4	None	4.1	<.1	6.61	*.252*	.02
	Minimal	1.7	<.1			
	Mild	8.0	.1			
	Moderate	5.9	.1			
	Mod. Sev	0	0			
	Severe	12.5	.1			

Note. EVI: Embedded validity indicator; None: PHQ-9 score of 0; Minimal: PHQ-9 score of 1-4; Mild PHQ-9 score of 5-9; Moderate: PHQ-9 score of 10-14; Mod. Sev: PHQ-9 score of 15-19; Severe: PHQ-9 score greater than 19 (Kroenke, Spitzer, & Williams, 2001); Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Table 18

Proportion Failing, Risk Ratio, and χ^2 analysis according to Depressive Symptoms reported at ≥1, ≥2, and ≥3 Conservative EVIs failed

# of EVI failures	Symptom endorsement	%	RR	χ^2	*p*	ϕ^2
≥1	None	36.7	.6	6.24	*.283*	.08
	Minimal	28.8	.4			
	Mild	41.4	.7			
	Moderate	37.3	.6			
	Mod. Sev	53.8	1.2			
	Severe	50	1			
≥2	None	12.2	.1	3.72	*.59*	.107
	Minimal	7.6	.1			
	Mild	16.1	.2			
	Moderate	11.8	.1			
	Mod. Sev	15.4	.2			
	Severe	12.5	.1			
≥3	None	2.0	>.1	5.90	*.316*	.02
	Minimal	2.5	>.1			
	Mild	6.9	.1			
	Moderate	7.8	.1			
	Mod. Sev	0	0			
	Severe	12.5	.1			

Note. None: PHQ-9 score of 0; Minimal: PHQ-9 score of 1-4; Mild PHQ-9 score of 5-9; Moderate: PHQ-9 score of 10-14; Mod. Sev: PHQ-9 score of 15-19; Severe: PHQ-9 score greater than 19 (Kroenke, Spitzer, & Williams, 2001); Conservative cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 7 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard test (GPB) dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Although raw scores provide insight into the frequency of depressive symptoms endorsed by participants, they are not indicative of symptom severity. Therefore, differences in the total number of EVIs failed was compared between four levels of symptom severity, as indicated by participants on Question 10 of the PHQ-9 (Table 19 and Table 20). Participants who endorsed no symptoms on questions 1-9 were excluded from the analysis. The level of symptom severity did not influence the number of total number of liberal or conservative EVIs failed.

Table 19

Comparison of the total number of Liberal EVIs failed across PHQ-9 Distress groups

	Total EVIs failed						
Level of Distress	n	*M*	SD	F	*df*	*p*	η_p^2
Not Very Difficult	123	1.0	1.2	.744	3	*.526*	.008
Somewhat Difficult	131	1.0	1.2				
Very Difficult	16	1.0	1.3				
Extremely Difficult	5	1.8	1.5				

Note. Level of distress as indicated by question 10 on the PHQ-9 (Kroenke, Spitzer, & Williams, 2001); Liberal cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 8 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 2 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 6 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 7 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard Test (GPB) dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 33 (Erdodi, Seke, et al., 2017); FAS ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 33 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

Table 20

Comparison of the total number of Conservative EVIs failed across PHQ-9 Distress groups

	Total EVIs failed						
Level of Distress	n	*M*	SD	F	*df*	*p*	η_p^2
Not Very Difficult	123	.46	.9	1.10	3	*.349*	.012
Somewhat Difficult	131	.60	.9				
Very Difficult	16	.75	1.2				
Extremely Difficult	5	1.0	1.3				

Note. Level of distress as indicated by question 10 on the PHQ-9 (Kroenke, Spitzer, & Williams, 2001); Conservative cutoffs: Hopkins Verbal Learning Test – Revised recognition discrimination (HVLT-R RD) ≤ 7 (Sawyer, Testa, & Dux, 2017); Wisconsin Card Sorting Test failure to maintain set (WCST FMS) ≥ 1 (King, Sweet, Sherer, Curtiss, & Vanderploeg, 2002; Larrabee 2003; Suhr & Boyer, 1999); Delis-Kaplan Executive Functioning System (D-KEFS) Trail 2 ≤ 5 (Erdodi, Hurtubise, et al., 2018); D-KEFS Trail 4 ≤ 6 (Erdodi, Hurtubise, et al., 2018); Grooved Pegboard test (GPB) dominant hand ≤ 29 (Erdodi, Seke, et al., 2017); GPB non-dominant hand ≤ 31 (Erdodi, Seke, et al., 2017); FAS ≤ 31 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015); Animals ≤ 29 (Curtis, Thompson, Greve, & Biachini, 2008; Sugarman & Axelrod, 2015).

CHAPTER IV

Discussion

Performance validity testing has been identified as a critical component of neuropsychological assessment (Board of Directors, 2007; Bush et al., 2005). However, EVI utility in an HIV+ population is yet to be explored. The objective of the current study was to begin investigating performance validity within a traditional neuropsychological HAND battery. To do this, the utility of several EVIs were examined within an undergraduate sample. Cumulative EVI failure was compared across six groups (e.g. controls, EXP_{MAL}, NN, ANI, MND, and HAD). The relationship between EVI failure and neurocognitive performance as well as self-reported depression was explored.

Within the present HAND battery, eight EVIs were identified that demonstrated acceptable signal detection profiles (AUC $\geq .70$) within the student sample. Consistent with our *a priori* hypothesis, the number of EVIs failed differed across the six groups, and was predictive of neuropsychological test scores within the HIV+ sample. Interestingly, and contrary to expectations, depression as measured by the PHQ-9 was independent of EVI failure.

Determining EVI cutoffs using the student sample

The signal detection profiles of eight EVIs were explored prior to their application to the HIV+ sample. All individual EVIs were significant predictors of the criterion variable (experimental malingering or Pass/Fail status on the WCT),and produced a classification accuracy hovering around the *Larrabee limit*: .50 sensitivity at .90 specificity (Erdodi et al., 2014). Failure of a single EVI is insufficient evidence of invalid performance (Boone, 2013; Rickards, Cranston, Touradji, & Bechtold, 2017). Therefore, participants were classified based on their cumulative EVI failures as *valid* (≤2 failures at liberal cutoffs; ≤1 failures at

conservative cutoffs), *borderline* (3 failures at liberal cutoffs; 2 failures at conservative cutoffs), or *invalid* (≥4 failures at liberal cutoffs; ≥3 failures at conservative cutoffs). The "indeterminate range" (i.e., *borderline*) has been recently introduced as a third category in the traditional binary classification system to formally acknowledge the inherent uncertainty in the performance validity assessment (Erdodi, 2017).

The present study supports the dual criterion model of EVI calibration, which suggests that multiple criterion PVTs should be used when calibrating an EVI to monitor potential instrumentation artifacts (Erdodi, Hurtubise, et al., 2018). Previous research suggested that the sensory modality and cognitive domain of a criterion PVT may inflate the sensitivity and specificity profile of an EVI with congruent features. Conversely, incongruence may lead to underestimating the classification accuracy (Erdodi, 2017; Erdodi & Roth, 2017).

Indeed, the effect of domain specificity was observed in the current study. For example, the WCT appears to be a measure of word list learning and memory and has been validated as a free-standing PVT (Barhon, Batchelor, Meares, Chekaluk, & Shores, 2015; Davis, 2014; Erdodi, Kirsch, Lajiness-O'Neill, Vingilis, & Medoff, 2014; Erdodi et al., 2017; Pearson, 2009). Within the present study, the WCT had a strong relationship with HVLT-R, a measure of auditory verbal learning, memory, and recognition, and weaker relationships with D-KEFS Trails, a measure of visuomotor processing speed and cognitive flexibility. It is worth noting that a strong relationship was observed between the WCT and GPB-ND (a test of manual dexterity), which suggests that the domain/modality specificity effect may be more complex than initially proposed, or that the WCT is a robust instrument that provides an unbiased index of performance validity within the HAND battery.

The experimental malingering paradigm was used as an alternative criterion variable, as it theoretically circumvents this limitation by allowing examinees to decide on which tests they choose to "demonstrate impairment". At the same time, the most notable weakness of experimental malingering as a criterion is the absence of any real external incentive to perform poorly while avoiding detection. As such, studies relying on this design have been criticized for inflating the classification accuracy of predictor PVTs by creating diagnostically pure groups with minimal overlap but little etiological validity. In this case, the signal detection model is applied to a measurement context that is unrealistically easy to characterize, thereby reducing its ecological validity. Indeed, cutoffs developed using the experimental malingering paradigm often fail to replicate (Rogers, 2008).

Similarly, experimental malingering comes with unique threats to internal validity. Namely, researchers rely exclusively on participants' ability and willingness to adhere to instructions. In reality, participants assigned to the control group often fail PVTs (An et al., 2012; 2017; 2018), contaminating the *valid* group. Conversely, participants assigned to the experimental malingering group may not make a genuine effort to produce credible impairment. Indeed, large variability in participants' execution of instructions was observed within the present study: the total number of EVI failures ranged from 0 to 8 regardless of whether liberal or conservative cutoffs were used. In other words, some made little-to-no attempt to appear impaired, while others excessively exaggerated deficits.

In contrast, using the WCT provides an opportunity to operationalize valid versus invalid responding using a well-established instrument. Psychometric definitions of non-credible responding have the advantage of refraining from making specific (and often untestable) assumptions about the validity of a given neurocognitive profile. It also allows researchers to

correct for the shortcomings of the experimental malingering design, by psychometrically defining *valid* and *invalid*. Thus, it correctly reclassifies examinees who were supposed to malinger but didn't and those who were supposed to perform at true ability but didn't as *valid* and *invalid*, respectively.

Hypothesis 1: EVI failure will vary according to group membership (EXP_{MAL}= HAD > MND = ANI > NN = CON).

As hypothesized, between-group differences in cumulative EVI failure were observed. Controls and NN individuals had the fewest number of EVI failures, supporting the notion that EVI failure is insensitive to variations in education, age, and gender (An et al., 2012). On average, both controls and NN had cumulative EVI profiles that were classified as *valid*. HAD and EXP_{MAL} groups failed the largest number of EVIs and produced invalid profiles at both liberal and conservative cutoffs. As mentioned, certain clinical populations, such as dementia, are exempt from PVTs due to genuine and severe neurological impairments that provide a more clinically accurate interpretation for a high number of PVT failures (Boone, 2013; Merten, Bossink, & Schmand, 2007).

Although individuals diagnosed with ANI had a greater number of EVI failures than controls, they produced valid profiles, supporting the use of EVIs in HAND, as the validity cutoffs do not misclassify individuals with genuine mild cognitive deficits as non-credible. Individuals with MND had a greater number of EVI failures than those with ANI, and on average, produced valid profiles when liberal cutoffs were applied but borderline profiles when applying conservative cutoffs. These findings suggest that individuals with MND can be effectively protected against being misclassified as non-credible using more liberal multivariate cutoffs, but they are vulnerable to false positive errors if conservative cutoffs are applied.

Nevertheless, the elevated mean number of EVI failures in this group suggests that the neurocognitive profiles of individuals with MND are at a higher risk for being misclassified as *invalid*. Therefore, the issue of false positives in this population warrants further investigation.

Differentiating invalid responding from genuine impairment is beyond the scope of the present study. However, a strong linear relationship between the severity of cognitive impairment and EVI failure was observed within the HIV+ sample (HAD > MND > ANI > NN; Antinori et al., 2007). Although the true nature and clinical interpretation of this dose-response relationship remains unclear, there are three potential explanations.

First, as EVIs are contained within ability tests, they are more prone to false positives in individuals with genuine cognitive impairments (Boone, 2013; DeRight & Carone, 2015). Therefore, the increasing number of EVI failures with HAND severity may reflect an elevated rate of false positives. Alternatively, the criteria used to determine an individual's HAND diagnosis may have been contaminated by non-credible responding, such that invalid performance resulted in the misclassification of an individual's cognitive profile as impaired. Lastly, rather than due to cognitive performance, the relationship between HAND severity and EVI failure may result from a secondary variable (e.g. pain, fatigue, depression, complex trauma history) that is commonly comorbid with HAND severity and accompanies increased risk of EVI failure (Bigler, 2014).

Hypothesis 2: Those failing a greater number of EVIs will perform more poorly on neuropsychological tests.

A persistent concern in performance validity assessment is that elevated BR_{Fail} is a consequence of false positives in neurocognitively impaired individuals (Bigler, 2014). To

investigate this idea, only individuals diagnosed as NN, ANI, and MND were included within the analysis.

Overall, individuals with valid profiles outperformed those with invalid profiles on six measures of cognitive ability. The cognitive profiles of the *borderline* cases tended to present more similarly to *invalid*, rather than *valid*, profiles, as reported previously (Erdodi, 2017). Taken together, the findings suggest that an inverse relationship exists between neurocognitive performance and EVI failure, replicating previous studies in undergraduates (An, Zakzanis, and Joordens, 2012; DeRight and Jorgensen, 2015), mixed clinical samples (Erdodi, Abeare, et al., 2017), and non-litigating epilepsy surgery candidates (Loring, Lee, & Meador, 2004).

EVIs are contained within neuropsychological tests. Therefore, collinearity (i.e., shared error variance) cannot be fully eliminated, as the same test is used to measure both cognitive ability and performance validity. Within the present study, all variables that were used as validity indicators were excluded from this analysis. The limited number of tests within the HAND battery required some cognitive measures to be derived from other variables within a given test (e.g. verbal fluency categories, HVLT-R total learning, HVLT-R delayed recall, and WCST categories). Within these cognitive measures, the relationship between cognitive ability and EVI failure may also be influenced by collinearity. However, the WRAT-4 reading subtest and SDMT did not contain any EVIs. Although it is impossible to completely separate different cognitive domains, the WRAT-4 reading subtest emphasizes a domain (e.g. reading) not contained within the other neuropsychological tests within the battery. Valid profiles outperformed borderline and invalid profiles on the WRAT-4 and SDMT, suggesting that individuals who fail a greater number of EVIs are more likely to have impaired scores in general,

rather than in a domain, modality, or test-specific manner, reinforcing the global deleterious effect of non-credible responding (Boone, 2013; Larrabee, 2012).

Hypothesis 3: Regardless of HIV status, individuals who have elevated PHQ-9 scores will fail more EVIs than those who do not.

An alternate explanation for Hypothesis 1 is that a third variable, such as depression, may account for between-group differences in EVI failure. It was hypothesized that individuals reporting more symptoms of depression would fail a greater number of EVIs. This prediction was not supported by the data: no relationship was observed between PHQ-9 scores and EVI failure. Further, in those reporting elevated levels of depression, symptom severity was independent of total EVI failures.

Previous research suggested BR_{Fail} is related to emotional distress, somatic concerns, fatigue, pain, sensory disturbances, and limited English proficiency (Erdodi, Seke et al., 2017; Erdodi, Nussbaum, et al., 2017; Erdodi et al., 2016; Greher & Wodushek, 2017, Whiteside et al., 2010). Further, PVT failure has been linked to depression defined by a score of ≥19 on the Beck Depression Inventory-II (BDI-II; McCormick, Yoash-Gantz, McDonald, Campbell, & Tupler, 2013). However, similar to our study, An, Zakzanis, and Joordens (2012) found no difference between valid and invalid profiles on the BDI in a student sample. The inconsistent relationship between self-reported depression and PVT failure may reflect methodological differences. Clinical depression is often comorbid with neurological problems (Christopher & MacDonald, 2010), and can even manifest as cognitive deficits in individuals without neurological disorders. Indeed, certain defining features of depression (psychomotor retardation, low energy, diminished ability to think and concentrate, indecisiveness; American Psychiatric Association, 2013) can manifest as impairment on neuropsychological testing (i.e., low visuomotor speed and slow

simple reaction time, attention, working memory). Therefore, when comparing PVT failures between those with probable depression and those without as done by McCormick and colleagues (2013), genuine cognitive impairments may inflate BR_{Fail}. In contrast, the present study and that conducted by An and colleagues (2012) investigated PVT failure across a continuum of depressive symptoms and may not have captured clinical impairments related to depression.

Limitations

Several limitations within the present study reduce the generalizability of its findings. The most notable is the use of an HIV- undergraduate sample to determine the signal detection profile and cutoff scores of EVIs in an HIV+ population. In contrast to the HIV+ sample, undergraduates were younger, more educated, and comprised of a greater proportion of females. Although theoretically resistant to their influences, the demographic variables of the student sample may have produced inappropriately liberal cutoffs when applied to the HIV+ sample. However, as neurocognitive tests and EVI cutoffs primarily utilize scales that correct for age, education, and gender it is unlikely that such factors (e.g. cognitive reserve) would influence the findings of the present study. An additional demographic variable not taken into consideration is limited English proficiency. As English proficiency increases the risk of EVI failure, undergraduate and HIV+ group differences may reduce the translatability of the EVIs (Erdodi, Nussbaum, et al., 2017). A related limitation is the lack of free-standing PVTs included within the HAND battery. Free-standing PVTs are designed to estimate response credibility and are much less sensitive to genuine impairment than EVIs (Greher & Wodushek, 2017). Their inclusion within the HAND battery would provide several PVT criteria to determine sensitivity

and specificity of EVI cutoffs within the HIV+ sample. Further, free-standing PVTs would provide EVI-independent insight into the credibility of a response set on a case-by-case basis.

A general limitation to HAND research is the way HAND status is assigned. Functional impairments are determined via self-report measures. These measures ultimately distinguish between ANI and MND groups (Antinori et al., 2007). Without an objective measure of functional changes, patient descriptions of subjective experiences may create artificial group differences.

Future Directions

The present study addressed limitations of previous PVT research using a single-blind paradigm in the undergraduate sample (An, Zakzanis, & Joordens, 2012) and relying on multiple criterion PVT. However, the current study is not without limitation. The data collected from the HIV+ population were archival and included no criterion PVT. Therefore, the signal detection analyses for EVIs within the HAND battery could not be extended to the HIV+ population. Future studies investigating cognitive functioning within the HAND population should include free-standing PVTs to allow researchers to address the collinearity issue (i.e., the confluence of genuine impairment and non-credible responding) and improve the internal validity of signal detection analyses.

The relationship between neurocognitive performance and EVI failure was confounded by collinearity in all but two tests. By including several independent tests encompassing a variety of cognitive domains, future studies could explore whether EVI failure and cognitive impairments occur in a domain-specific or random pattern. If the EVIs failed are restricted to tests assessing performance within a single cognitive domain, then it is likely that non-credible profiles are a consequence of inflated false positive rates. Thus, future research in this area

would be supported by validating EVIs using HIV+ samples and increasing the number of neurocognitive tests included within the battery. In the meantime, for clinical purposes, at least two free-standing PVTs should be routinely administered in addition to the standard HAND battery, consistent with the recommendations of professional organizations (Bush et al., 2005; Heilbronner et al., 2009).

Summary and Conclusions

The present study is the first to demonstrate that BR_{Fail} varies on EVIs within the HAND battery as a function of classification severity in an HIV+ population. These results may reflect elevated false positives due to genuine cognitive impairment or due to a third unmeasured variable. Worth mentioning, although the number of EVIs failed differed between groups, the average individual from NN and ANI groups was classified as having a valid profile. Although EVIs have been validated as a means to determine non-credible responding in several clinical populations, they are considered inappropriate in others (i.e. intellectual disability and dementia; Davis, 2018; Shandera et al., 2010). Thus, HAD populations should be excempt from performance validity testing. It is too early to determine whether MND due to HIV+ should be added to the list of exempt categories, but it warrants further empirical research.

Even in populations where EVI use is recommended, other factors besides invalid performance have been identified as contributing to EVI failure including alterations in white matter integrity (Clark et al., 2016), pain, and sleep (Johnson-Greene, Brooks, & Ference, 2013). All three of these factors are prevalent in HIV+ populations and may contribute to increased risk of EVI failure. The present study suggests that it is unlikely that the factor underlying increased EVI failure is depression.

In contrast, the relationship between EVI failure and HAND diagnosis may be due to invalid performance resulting in lower cognitive test scores and a more severe clinical diagnosis. Over a one year period, cognitive impairment levels fluctuate in almost half of individuals diagnosed with HAND (Antinori et al., 2007). If invalid responding is the true cause of low scores (rather than genuine impairment), it could provide an explanation for these fluctuations in the cognitive profile. The general relationship between cognitive impairment and EVI failure may support this hypothesis but further research is required, given the clinical implications of the conclusion (fluctuations of test taking effort vs. cognitive ability).

REFERENCES

Akhtar-Khaleel, W. Z., Cook, R. L., Shoptaw, S., Miller, E. N., Sacktor, N., Surkan, P. J., … Plankey, M. (2017). Association of midlife smoking status with change in processing speed and mental flexibility among HIV-seropositive and HIV-seronegative older men: the Multicenter AIDS Cohort Study. *Journal of NeuroVirology*, *23*(2), 239–249. doi: 10.1007/s13365-016-0496-6

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). Washington, D.C: American Psychiatric Association.

An, K. Y., Charles, J., Ali, S., Enache, A., Dhuga, J., & Erdodi, L. A. (2018). Reexamining performance validity cutoffs within the Complex Ideational Material and the Boston Naming Test-Short Form using an experimental malingering paradigm. *Journal of Clinical and Experimental Neuropsychology.* Advanced online publication. doi: 10.1080/13803395.2018.1483488.

An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2017). Performance validity in undergraduate research participants: a comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist*, *31*(1), 193–206. doi: 10.1080/13854046.2016.1217046

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*(8), 849-857. doi: 10.1093/arclin/acs085.

Antinori, A., Arendt, G., Becker, J. T., Brew, B. J., Byrd, D. A., Cherner, M., … Wojna, V. E. (2007). Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*, *69*(18), 1789–1799. doi:10.1212/01.WNL.0000287431.88658.8b

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist, 21*(6), 899–916. doi: /10.1080/13825580600966391

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various Digit Span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145–159. doi: 10.1080/13854040590947362

Barhon, L. I., Batchelor, J., Meares, S., Chekaluk, E., & Shores, E. A. (2015). A comparison of the degree of effort involved in the TOMM and the ACS Word Choice Test using a dual-task paradigm. *Applied Neuropsychology: Adult, 22,* 114-123.

Basso, M. R., & Bornstein, R. A. (2003). Effects of past noninjection drug abuse upon executive function and working memory in HIV Infection. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A), 25*(7), 893–903. doi: 10.1076/jcen.25.7.893.16489

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF Manual for Administration, Scoring, and Interpretation.* Minneapolis: University of Minnesota Press.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*(7-8), 909–916. doi: 10.3109/02699052.2013.793400

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*(13-14), 1623–1638. doi: 10.3109/02699052.2014.947627

Bilder, R. M., Sugar, C. A., & Hellemann, G. S. (2014). Cumulative false positive rates given multiple performance validity tests: Commentary on Davis and Millis (2014) and

Larrabee (2014). *The Clinical Neuropsychologist*, *28*(8), 1212–1223. doi: 10.1080/13854046.2014.969774

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology*, *23*(4), 379–391. doi: 10.1016/j.acn.2008.01.008

Board of Directors. (2007). American Academy of Clinical Neuropsychology (AACN) Practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, *21*(2), 209–231. doi: 10.1080/13825580601025932

Boone, K. B. (2013). *Clinical Practice of Forensic Neuropsychology*. New York, NY: Guilford.

Boone, K. B., & Lu, P. (2003). Noncredible cognitive performance in the context of severe brain injury. *Clinical Neuropsychology, 17*(2), 244-254. Doi:10.1076/clin.16.2.244.16497

Brandt, J., & Benedict, R. H. (2001). *Hopkins Verbal Learning Test— Revised.* Lutz, FL: Psychological Assessment Resources, Inc.

Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pliskin, N., … Silver, C. (2005). Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. doi: 10.1016/j.acn.2005.02.002

Busse, M., & Whiteside, D. (2012). Detecting suboptimal cognitive effort: Classification accuracy of the Conner's Continuous Performance Test-II, Brief Test of Attention, and Trail Making Test. *The Clinical Neuropsychologist*, *26*(4), 675–687. doi: 10.1080/13854046.2012.679623

Carey, C., Woods, S., Rippeth, J., Gonzalez, R., Moore, D., Marcotte, T., … HNRC Group. (2004). Initial validation of a screening battery for the detection of HIV-associated cognitive impairment. *The Clinical Neuropsychologist, 18*(2), 234–248. doi: 10.1080/13854040490501448

Carone, D. A. (2015). Assessment of response bias in neurocognitive evaluations. *NeuroRehabilitation, 36*(4), 387–400. doi: 10.3233/NRE-151228

Chang, L., Lim, A., Lau, E., & Alicata, D. (2017). Chronic tobacco-smoking on psychopathological symptoms, impulsivity and cognitive deficits in HIV-infected individuals. *Journal of Neuroimmune Pharmacology*. doi:10.1007/s11481-017-9728-7

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. *Cognitive Neuropsychiatry, 10*(5), 379-399. Doi: 10.1080/13546800444000128.

Clark, A. L., Sorg, S. F., Schiehser, D. M., Bigler, E. D., Jacobson, M. W., … Delano-Wood, L. (2016). White matter associations with performance validity testing in veterans with mild traumatic brain injury: The utility of biomarkers in complicated assessment. *Journal of Head Trauma Rehabilitation, 31*(5), 246-259. doi: 10.1097/HTR.0000000000000183

Cohen, R. A., de la Monte, S., Gongvatana, A., Ombao, H., Gonzalez, B., Devlin, K. N., … Tashima, K. T. (2011). Plasma cytokine concentrations associated with HIV/hepatitis C coinfection are related to attention, executive and psychomotor functioning. *Journal of Neuroimmunology, 233*(1-2), 204–210. doi: 10.1016/j.jneuroim.2010.11.006

Cole, M. A., Margolick, J. B., Cox, C., Li, X., Selnes, O. A., Martin, E. M., […] Miller, E. N. (2007). Longitudinally preserved psychomotor performance in long-term asymptomatic HIV-infected individuals. *Neurology, 69*(24), 2213–2220. doi: 10.1212/01.WNL.0000277520.94788.82

Corrêa, D. G., Zimmermann, N., Netto, T. M., Tukamoto, G., Ventura, N., de Castro Bellini Leite, S., … Gasparetto, E. L. (2016). Regional cerebral gray matter volume in HIV-positive patients with executive function deficits: Cerebral gray matter volume in HIV-positive patients. *Journal of Neuroimaging*, *26*(4), 450–457. doi: 10.1111/jon.12327

Curtis, K. L., Thompson, L. K., Greve, K. W., & Bianchini, K. J. (2008). Verbal fluency indicators of malingering in traumatic brain injury: Classification accuracy in known groups. *The Clinical Neuropsychologist*, *22*(5), 930–945. doi: 10.1080/13854040701563591

Cysique, L. A., Franklin, D., Abramson, I., Ellis, R. J., Letendre, S., Collier, A., … the HNRC group. (2011). Normative data and validation of a regression based summary score for assessing meaningful neuropsychological change. *Journal of Clinical and Experimental Neuropsychology*, *33*(5), 505–522. doi: 10.1080/13803395.2010.535504

Cysique, L., Maruff, P., Darby, D., & Brew, B. (2006). The assessment of cognitive function in advanced HIV-1 infection and AIDS dementia complex using a new computerised cognitive test battery. *Archives of Clinical Neuropsychology*, *21*(2), 185–194. doi: 10.1016/j.acn.2005.07.011

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. H. M. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, *39*(3), 257–264. doi: 10.1080/13803395.2016.1223278

Davis, J. J. (2014). Further consideration of Advanced Clinical Solutions Word Choice: Comparison to the Recognition Memory Test – Words and classification accuracy on a clinical sample. *The Clinical Neuropsychologist, 28*(8), 1278-1294. doi:10.1080/13854046.2014.975844

Davis, J. J. (2018). Performance validity in older adults: observed versus predicted false positive rates in relation to number of tests administered. *Journal of Experimental Neuropsychology*. Advanced online publication. Doi: 10.1080/13803395.2018.1472221

de Almeida, M., Kamat, R., Cherner, M., Umlauf, A., Ribeiro, C. E., de Pereira, A. P., … Ellis, R. J. (2017). Improving detection of HIV-associated cognitive impairment: Comparison of the international HIV Dementia Scale and a Brief Screening Battery. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *74*(3), 332–338. doi: 10.1097/QAI.0000000000001224

Dean, P. J. A., & Sterr, A. (2013). Long-term effects of mild traumatic brain injury on cognitive performance. *Frontiers in Human Neuroscience, 7*, 30. doi: 10.3389/fnhum.2013.00030

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.

DenBoer, J. W., & Hall, S. (2007). Neuropsychological test performance of successful brain injury simulators. *The Clinical Neuropsychologist*, *21*(6), 943-955.

DeRight, J., & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology*, *21*(1), 1–24. doi: 10.1080/09297049.2013.864383

DeRight, J., & Jorgensen, R. S. (2015). I just want my research credit: frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *Clinical Issues, 29*(1), 101-117. doi: 10.1080/13854046.2014.989267

Eggers, C., Arendt, G., Hahn, K., Husstedt, I. W., Maschke, M., Neuen-Jacob, E., … Straube, E. (2017). HIV-1-associated neurocognitive disorder: epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Neurology*. doi: 10.1007/s00415-017-8503-2

Ellis, R. J., Deutsch, R., Heaton, R. K., Marcotte, T. D., McCutchan, J. A., Nelson, J. A., … San Diego HIV Neurobehavioral Research Center Group. (1997). Neurocognitive impairment is an independent risk factor for death in HIV infection. *Archives of Neurology, 54*(4), 416–424. doi:10.1001/archneur.1997.00550160054016

Erdodi, L. A. (2017) Aggregating validity indicators: The salience of domain specificity and the indeterminant range in multivariate models of performance validity assessment. *Applied Neuropsychology: Adult.* Advanced online publication. doi: 10.1080/23279095.2017.1384925.

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment, 29*(2), 148–157. doi: 10.1037/pas0000319

Erdodi, L. A., Kirsch, N. L., Lajiness-O'Neill, R., Vingilis, E., & Medoff, B. (2014). Comparing the Recognition Memory Test and the Word Choice Test in a mixed clinical sample: Are they equivalent? *Psychological Injury and Law, 7*(3), 255-263.

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: an emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist, 31*(6-7), 1029-1046. Doi: 10.1080/13854046.2017.1323119.

Erdodi, L. A., Nussbaum, S., Sagar, S., Abeare, C. A., & Schwartz, E. S. (2017). Limited english proficiency increases failure rates on performance validity tests with high verbal mediation. *Psychological Injury and Law, 10*(1), 96–103. doi: 10.1007/s12207-017-9282-x

Erdodi, L., & Roth, R. (2016). Low scores on BDAE Complex Ideational Material are associated with invalid performance in adults without aphasia. *Applied Neuropsychology: Adult*, http://dx.doi.org/10.1080/23279095.2016.1154856.

Erdodi, L. A., Seke, K., Shahein, A., Tyson, B. T., Sagar, S., & Roth, R. (2017). Low scores on the Grooved Pegboard Test are associated with invalid responding and psychiatric symptoms. *Psychology and Neuroscience.*

Erdodi, L. A., Tyson, B. T., Abeare, C. A., Lichtenstein, J. D., Pelletier, C. L., Rai, J. K., & Roth, R. M. (2016). The BDAE Complex Ideational Material—A measure of receptive language or performance validity? *Psychological Injury and Law*, *9*(2), 112-120. doi:http://dx.doi.org.ezproxy.lib.ryerson.ca/10.1007/s12207-016-9254-6

Erdodi, L. A., Tyson, B. T., Shahein, A. G., Lichtenstein, J. D., Abeare, C. A., Pelletier, C. L., ... & Roth, R. M. (2016). The power of timing: Adding a time-to-completion cutoff to the Word Choice Test and Recognition Memory Test improves classification accuracy. *Journal of Clinical and Experimental Neuropsychology*, 1-15.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988a). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, *19*(5), 508–515.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988b). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing, *Journal of Consulting and Clinical Psychology, 56*(4), 578–582.

Fellows, R. P., Byrd, D. A., Elliott, K., Robinson-Papp, J., Mindt, M. R., & Morgello, S. (2012). Distal sensory polyneuropathy is associated with neuropsychological test performance among persons with HIV. *Journal of the International Neuropsychological Society*, *18*(05), 898–907. doi: 10.1017/S1355617712000707

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury*, *21*(4), 373–383. doi: 10.1080/02699050701311133

Frederick, R. I. (2003). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology*, *2*(3-4), 1–25. doi: 10.1300/J151v02n03_01

Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2018). Examining false-positive rates of Wechsler Adult Intelligence Scale (WAIS-IV) processing speed-based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment.* Advanced online publication. doi: 10.1037/pas0000650.

Gomez, D., Power, C., Gill, M. J., & Fujiwara, E. (2017). Determinants of risk-taking in HIV-associated neurocognitive disorders. *Neuropsychology, 31*(7), 798-810. doi: 10.1037/neu0000366.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen III, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. doi: 10.1080/02699050110088254

Greher, M. R., & Wodushek, T. R. (2017). Performance validity testing in neuropsychology: Scientific basis and clinical application—A Brief Review. *Journal of Psychiatric Practice*, *23*(2), 134–140. doi: 10.1097/PRA.0000000000000218

Greiffenstein, M., Gervais, R., Baker, W. J., Artiola, L., & Smith, H. (2013). Symptom validity testing in medically unexplained pain: A chronic regional pain syndrome type 1 case series. *The Clinical Neuropsychologist*, *27*(1), 138–147. doi: 10.1080/13854046.2012.722686

Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction with the Wisconsin Card Sorting Test: A preliminary investigation in traumatic brain injury. *The Clinical Neuropsychologist, 16*(2), 179–191. doi: 10.1076/clin.16.2.179.13241

Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, *90*(7), 1117–1126. doi: 10.1016/j.apmr.2009.01.018

Gustafson, D. R., Mielke, M. M., Tien, P. C., Valcour, V., Cohen, M., Anastos, K., … Crystal, H. A. (2013). Anthropometric measures and cognition in middle-aged HIV-infected and uninfected women. The Women's Interagency HIV Study. *Journal of NeuroVirology, 19*(6), 574–585. doi: 10.1007/s13365-013-0219-1

Heaton, R. K., Clifford, D. B., Franklin, D. R., Woods, S. P., Ake, C., Vaida, F., […]For the CHARTER Group. (2010). HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. *Neurology, 75*(23), 2087–2096. doi: 10.1212/WNL.0b013e318200d727

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). Wisconsin Card Sorting Test (WCST) manual revised and expanded. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Franklin, D. R., Ellis, R. J., McCutchan, J. A., Letendre, S. L., LeBlanc, S., … Grant, I. (2011). HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. *Journal of NeuroVirology*, *17*(1), 3–16. doi: 10.1007/s13365-010-0006-1

Heaton, R. K., Miller, S. W., Taylor, M. K., & Grant, I. (2004). *Revised comprehensive norms for an Expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources.

Heaton, R. K., Marcotte, T. D., Mindt, M. R., Sadek, J., Moore, D. J., Bentley, H., … Grant, I. (2004). The impact of HIV-associated neuropsychological impairment on everyday functioning. *Journal of the International Neuropsychological Society*, *10*(03). doi: 10.1017/S1355617704102130

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*(5), 892–900. doi: 10.1037/0022-006X.46.5.892

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

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (second edition).* New York, NY: John Wiley & Sons, Inc.

Erdodi, L. A, Hurtubise, J. L., Charron, C., Dunn, A. G., Enache, A., McDermott, A., Hirst, R. B. (2018). The D-KEFS trails as performance validity tests. *Psychological Assessment, 20*(8), 1082-1095. doi: 10.1037/pas000056

Iudicello, J. E., Woods, S. P., Parsons, T. D., Moran, L. M., Carey, C. L., & Grant, I. (2007). Verbal fluency in HIV infection: A meta-analytic review. *Journal of the International Neuropsychological Society*, *13*(01). doi: 10.1017/S1355617707070221

Iudicello, J. E., Woods, S. P., Weber, E., Dawson, M. S., Scott, J. C., Carey, C. L., … The HIV Neurobehavioral Research Ce. (2008). Cognitive mechanisms of switching in HIV-associated category fluency deficits. *Journal of Clinical and Experimental Neuropsychology*, *30*(7), 797–804. doi: 10.1080/13803390701779578

Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1994). Application of a forced-choice memory procedure designed to detect experimental malingering. *Archives of Clinical Neuropsychology, 9*(5), 437-450.

Iverson, G. L., Lange, R. T., Green, P., & Franzen, M. D. (2002). Detecting exaggeration and malingering with the Trail Making Test. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *16*(3), 398–406. doi: 10.1076/clin.16.3.398.13861

Jaggers, J. R., Dudgeon, W. D., Burgess, S., Phillips, K. D., Blair, S. N., & Hand, G. A. (2014). Psychological correlates of HIV-related symptom distress. *Journal of the Association of Nurses in AIDS Care*, *25*(4), 309–317. doi: 10.1016/j.jana.2013.06.003

Janssen, R. S., Cornblath, D. R., Epstein, L. G., Foa, R. P., McArthur, J. C., Price, R. W., & [...] Zaugid, M. (1991). Nomenclature and research case definitions for neurologic manifestations of human immunodeficiency virus-type 1 (HIV-1) infection. Report of a working group of the American Academy of Neurology AIDS Task Force. *Neurology*, *41*(6), 778–785.

Jasinski, L. J., Harp, J. P., Berry, D. T. R., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist*, *25*(8), 1415–1428. doi: 10.1080/13854046.2011.630024

Jodzio, K., & Biechowska, D. (2010). Wisconsin Card Sorting Test as a measure of executive function impairments in stroke patients. Applied Neuropsychology, 17(4), 267–277. doi:10.1080/09084282.2010.525104

Johnson-Greene, D., Brooks, L., & Ference, T. (2013). Relationship between performance validity testing, disability status, and somatic complaints in patients with fibromyalgia. *Clinical Neuropsychology. 27*(1), 148-158. Doi: 10.1080/13854046.2012.733732.

Jung, B., & Reidenberg, M. M. (2007). Physicians being deceived. *Pain Medicine*, *8*(5), 433–437. doi: 10.1111/j.1526-4637.2007.00315.x

Kamminga, J., Lal, L., Wright, E. J., Bloch, M., Brew, B. J., & Cysique, L. A. (2017). Monitoring HIV-associated neurocognitive disorder using screenings: A critical review including guidelines for clinical and research use. *Current HIV/AIDS Reports*, *14*(3), 83–92. doi: 10.1007/s11904-017-0349-9

Keen, L., & Turner, A. D. (2014). Association between interleukin-6 and neurocognitive performance as a function of self-reported lifetime marijuana use in a community based

sample of African American adults. *Journal of the International Neuropsychological Society, 20*(08), 773–783. doi: 10.1017/S1355617714000691

King, J. H., Sweet, J. J., Sherer, M., Curtiss, G., & Vanderploeg, R. D. (2002). Validity indicators within the Wisconsin Card Sorting Test: Application of new and previously researched multivariate procedures in multiple traumatic brain injury samples. The *Clinical Neuropsychologist, 16*(4), 506–523. doi:10.1076/clin.16.4.506.13912

Kinuthia, R. N., Thigiti, J. M., & Gakinya, B. N. (2016). Relationship between HIV stage and psychomotor speed neurocognitive score at a Kenyan sub-county hospital. *African Journal of Primary Health Care & Family Medicine, 8*(1). doi: 10.4102/phcfm.v8i1.1061

Kireev, M., Korotkov, A., Medvedeva, N., & Medvedev, S. (2013). Possible role of an error detection mechanism in brain processing of deception: PET-fMRI study. *International Journal of Psychophysiology*, *90*(3), 291–299. doi: 10.1016/j.ijpsycho.2013.09.005

Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Winsconsin Card Sorting-64 Card Version*. Odessa, FL: Psychological Assessment Resources, Inc.

Kore, I., Ananworanich, J., Valcour, V., Fletcher, J. L. K., Chalermchai, T., Paul, R., … Spudich, S. (2015). Neuropsychological impairment in acute HIV and the effect of immediate antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *70*(4), 393–399. doi: 10.1097/QAI.0000000000000746

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Ku, N., Lee, Y., Ahn, J., Song, J., Kim, M., Kim, S., … Choi, J. (2014). HIV-associated neurocognitive disorder in HIV-infected Koreans: the Korean NeuroAIDS Project: HAND in HIV-infected Koreans. *HIV Medicine, 15*(8), 470–477. doi: 10.1111/hiv.12137

Küper, M., Rabe, K., Esser, S., Gizewski, E. R., Husstedt, I. W., Maschke, M., & Obermann, M. (2011). Structural gray and white matter changes in patients with HIV. *Journal of Neurology, 258*(6), 1066–1075. doi: 10.1007/s00415-010-5883-y

Lake, J. E., Vo, Q. T., Jacobson, L. P., Sacktor, N., Miller, E. N., Post, W. S., … Brown, T. T. (2015). Adiponectin and interleukin-6, but not adipose tissue, are associated with worse neurocognitive function in HIV-infected men. *Antiviral Therapy, 20*(2), 235–244. doi: 10.3851/IMP2952

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 17*(3), 410–425. doi: 10.1076/clin.17.3.410.18089

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, *18*(4), 625–630.

Larrabee, G. J. (2014a). Minimizing false positive error with multiple performance validity tests: Response to Bilder, Sugar, and Hellemann (2014 this issue). *The Clinical Neuropsychologist*, *28*(8), 1230–1242. doi: 10.1080/13854046.2014.988754

Larrabee, G. J. (2014b). Test validity and performance validity: Considerations in providing a framework for development of an ability-focused neuropsychological test battery. *Archives of Clinical Neuropsychology*, *29*(7), 695–714. doi: 10.1093/arclin/acu049

Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury*, *24*(2), 89–99. doi: 10.3109/02699050903508218

Lawson, E., Sabin, C., Perry, N., Richardson, D., Gilleece, Y., Churchill, D., […] Walker-Bone, K. (2014). Is HIV painful? An epidemiologic study of the prevalence and risk factors for pain in HIV-infected patients. *The Clinical Journal of Pain*, *31*(9), 813-819. doi: 10.1097/AJP.0000000000000162

Leighton, A., Weinborn, M., & Maybery, M. (2014). Bridging the gap between neurocognitive processing theory and performance validity assessment among the cognitively impaired: a review and methodological approach. *Journal of International Neuropsychological Society, 20*(9), 873-886. Doi: 10.1017/S135561771400085X

Lichtenstein, J. D., Erdodi, L. A., Rai, J. K., Mazur-Mosiewicz, A., & Flaro, L. (2016). Wisconsin Card Sorting Test embedded validity indicators developed for adults can be extended to children. *Child Neuropsychology*, 1–14. doi: 10.1080/09297049.2016.1259402

Llorente, A. M., Brouwers, P., Leighty, R., Malee, K., Smith, R., Harris, L., … Chase, C. (2014). An analysis of select emerging executive skills in perinatally HIV-1-infected children. *Applied Neuropsychology: Child*, *3*(1), 10–25. doi: 10.1080/21622965.2012.686853

Loring, D. W., Lee, G. P., & Meador, K. M. (2004). Victoria Symptom Validity Test performance in non-litigating epilepsy surgery candidates. *Journal of Clinical and Experimental Neuropsychology. 27*(5). 610-617. Doi: 10.1080/13803390490918471

Mandal, N., Singh, O. P., Bhattacharya, S., Chatterji, S., Biswas, A., & Sen, S. (2008). Neurocognitive impairment in early HIV-positive individuals. *Journal of the Indian Medical Association, 106*(7), 442, 447–449, 453.

Manly, J. J., Smith, C., Crystal, H. A., Richardson, J., Golub, E. T., Greenblatt, R., … Young, M. (2011). Relationship of ethnicity, age, education, and reading level to speed and executive function among HIV+ and HIV– women: The Women's Interagency HIV Study (WIHS) Neurocognitive Substudy. *Journal of Clinical and Experimental Neuropsychology, 33*(8), 853–863. doi: 10.1080/13803395.2010.547662

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*(7), 1204–1237. doi: 10.1080/13854046.2010.514290

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist, 29*(6), 741–776. doi: 10.1080/13854046.2015.1087597

McCormick, C. K., Yoash-Gantz, R. E., McDonald, S. D., Campbell, T. C., & Tupler, L. A. (2013). Performance on the Green Word Memory Test following Operation Enduring Freedom/Operation Iraqi Freedom-era military service: Test failure is related to evaluation context. *Archives of Clinical Neuropsychology. 28*(8). 808-823. Doi: 10.1093/arclin/act050.

McIntosh, R. C., Rosselli, M., Uddin, L. Q., & Antoni, M. (2015). Neuropathological sequelae of Human Immunodeficiency Virus and apathy: A review of neuropsychological and neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *55*, 147–164.

doi: 10.1016/j.neubiorev.2015.04.008

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 308–318. doi:10.1080/ 13803390600693607

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*(1), 8–15. doi: 10.1093/arclin/acq083

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A), 24*(8), 1094–1102. doi: 10.1076/jcen.24.8.1094.8379

Mittenberg, W., Theroux-Fichera, S., Zielinski, R., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale--Revised. *Professional Psychology: Research and Practice*, *26*(5), 491–498. doi: 10.1037/0735-7028.26.5.491

Mohamed, M. A., Barker, P. B., Skolasky, R. L., Selnes, O. A., Moxley, R. T., Pomper, M. G., & Sacktor, N. C. (2010). Brain metabolism and cognitive impairment in HIV infection: a 3-T magnetic resonance spectroscopy study. *Magnetic Resonance Imaging, 28*(9), 1251–1257. doi: 10.1016/j.mri.2010.06.007

Monroe, A., Zhang, L., Jacobson, L., Plankey, M., Brown, T., Miller, E., … Sacktor, N. (2017). The association between physical activity and cognition in men with and without HIV infection. *HIV Medicine, 18*(8), 555-563. doi: 10.1111/hiv.12490

Moore, D. J., Roediger, M. J. P., Eberly, L. E., Blackstone, K., Hale, B., Weintrob, A., … Crum-Cianflone, N. F. (2012). Identification of an abbreviated test battery for detection of HIV-

associated neurocognitive impairment in an early-managed HIV-infected cohort. *PLoS ONE, 7*(11), e47310. doi: 10.1371/journal.pone.0047310

Moradi, A. R., Miraghaei, M. A., Parhon, H., Jabbari, H., & Jobson, L. (2012). Posttraumatic stress disorder, depression, executive functioning, and autobiographical remembering in individuals with HIV and in carers of those with HIV in Iran. *AIDS Care, 25*(3), 281–288. doi:10.1080/09540121.2012.701719

Morey, L.C. (2007). The Personality Assessment Inventory professional manual. Lutz, FL: Psychological Assessment Resources.

Muñoz-Moreno, J. A., Fumaz, C. R., Ferrer, M. J., Prats, A., Negredo, E., Garolera, M., … Clotet, B. (2008). Nadir CD4 cell count predicts neurocognitive impairment in HIV-infected patients. *AIDS Research and Human Retroviruses*, *24*(10), 1301–1307. doi: 10.1089/aid.2007.0310

Muñoz-Moreno, J. A., Prats, A., Pérez-Álvarez, N., Fumaz, C. R., Garolera, M., Doval, E., … Clotet, B. (2013). A brief and feasible paper-based method to screen for neurocognitive impairment in HIV-infected patients: The NEU screen. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *63*(5), 585–592. doi: 10.1097/QAI.0b013e31829e1408

Nolting, T., Lindecke, A., Hartung, H.-P., Koutsilieri, E., Maschke, M., Husstedt, I.-W., … Arendt, G. (2012). Cytokine levels in CSF and neuropsychological performance in HIV patients. *Journal of NeuroVirology*, *18*(3), 157–161. doi: 10.1007/s13365-012-0091-4

Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, *8*(1), 46–63. doi: 10.1007/s12207-015-9216-4

Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills*, *57*(2), 367–372. doi: 10.2466/pms.1983.57.2.367

Parker, R., Stein, D. J., & Jelsma, J. (2014). Pain in people living with HIV/AIDS: a systematic review. *Journal of the International AIDS Society*, *17*(1). doi: 10.7448/IAS.17.1.18719

Pereira, M., Fialho, R., & Canavarro, M. C. (2014). Prevalence and correlates of emotional distress in HIV/HCV coinfection. *AIDS Care*, *26*(sup1), S56–S64. doi: 10.1080/09540121.2014.906549

Pearson (2009). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV – Technical Manual*. San Antonio, TX: Author

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. Routledge.

Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaires de France.

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2017). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology: Adult*. Advanced online publication. doi: 10.1080/23279095.2017.1278602

Robertson, K., Jiang, H., Kumwenda, J., Supparatpinyo, K., Evans, S., Campbell, T. B., … the AIDS Clinical Trials Group. (2012). Improved neuropsychological and neurological functioning across three antiretroviral regimens in diverse resource-limited settings: AIDS clinical trials group study A5199, the international neurological study. *Clinical Infectious Diseases*, *55*(6), 868–876. doi: 10.1093/cid/cis507

A, R. (1988). *Clinical assessment of malingering and deception*. New York, NY: Guilford Press.

Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception*. Guilford Press.

Rubin, L. H., Cook, J. A., Weber, K. M., Cohen, M. H., Martin, E., Valcour, V., … Maki, P. M. (2015). The association of perceived stress and verbal memory is greater in HIV-infected versus HIV-uninfected women. *Journal of NeuroVirology*, *21*(4), 422–432. doi: 10.1007/s13365-015-0331-5

Rubinstein, M. L., & Selwyn, P. A. (1998). High prevalence of insomnia in an outpatient population with HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology: Official Publication of the International Retrovirology Association*, *19*(3), 260–265.

Ruffolo, L. F., Guilmette, T. J., & Willis, G. W. (2000). Comparison of time and error rates on the Trail Making Test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *14*(2), 223–230. doi: 10.1076/1385-4046(200005)14:2;1-Z;FT223

Sacktor, N., McDermott, M. P., Marder, K., Schifitto, G., Selnes, O. A., McArthur, J. C., … Epstein, L. (2002). HIV-associated cognitive impairment before and after the advent of combination therapy. *Journal of Neurovirology*, *8*(2), 136–142. doi: 10.1080/13550280290049615

Sacktor, N., Skolasky, R. L., Cox, C., Selnes, O., Becker, J. T., Cohen, B., … Miller, E. N. (2010). Longitudinal psychomotor speed performance in human immunodeficiency virus–seropositive individuals: impact of age and serostatus. *Journal of Neurovirology*, *16*(5), 335–341. doi: 10.3109/13550284.2010.504249

Sawyer, R. J., Testa, S. M., & Dux, M. (2017). Embedded performance validity tests within the Hopkins Verbal Learning Test – Revised and the Brief Visuospatial Memory Test – Revised. *The Clinical Neuropsychologist, 31*(1), 207–218. doi: 10.1080/13854046.2016.1245787

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, *19*(1), 21–30. doi: 10.1177/1073191111428764

Schutte, C., & Axelrod, B. N. (2013). Use of embedded cognitive symptom validity measures in mild traumatic brain injury cases. In D. A. Carone & S. S. Bush (Eds.). In *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 159–181). New York, NY: Springer.

Scott, J., Woods, S., Patterson, K., Morgan, E., Heaton, R., Grant, I., & Marcotte, T. (2006). Recency effects in HIV-associated dementia are characterized by deficient encoding☆. *Neuropsychologia, 44*(8), 1336–1343. doi: 10.1016/j.neuropsychologia.2006.01.008

Shandera, A. L., Berry, D. T.., Clark, J. A., Schipper, L. J.., Graue, L. O., & Harp, J. P. (2010). Detection of malingered mental retardation. *Pscyhological Assessment, 22*(1), 50-56. Doi: 10.1037/a0016585.

Sharland, M., & Gfeller, J. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213–223. doi: 10.1016/j.acn.2006.12.004

Shura, R. D., Miskey, H. M., Rowland, J. A., Yoash-Gantz, R. E., & Denning, J. H. (2016). Embedded performance validity measures with postdeployment veterans: Cross-

validation and efficiency with multiple measures. *Applied Neuropsychology: Adult*, *23*(2), 94–104. doi: 10.1080/23279095.2015.1014556

Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M.S. Sherman & B. L. Brooks (Eds.). In *Pediatric Forensic Neuropsychology* (pp. 113–135). New York, NY: Oxford University Press.

Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In *Mild traumatic brain injury: Symptom validity assessment and malingering. In D. A. Carone & S. S. Bush (Eds.)* (pp. 57–72). New York, NY: Springer.

Slick, D. J., Sherman, E. M. S., & Iverman. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *13*(4), 545–561. doi: 10.1076/1385-4046(199911)13:04;1-Y;FT545

Smith, A. (1973). Symbol Digit Modalities. Los Angeles, CA: Western Psychological Services.

Spies, G., Fennema-Notestine, C., Archibald, S. L., Cherner, M., & Seedat, S. (2012). Neurocognitive deficits in HIV-infected women and victims of childhood trauma. *AIDS Care*, *24*(9), 1126–1135. doi: 10.1080/09540121.2012.687813

Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using Verbal Fluency Tests in a clinical sample. *Applied Neuropsychology: Adult*, *22*(2), 141–146. doi: 10.1080/23279095.2013.873439

Suhr, J. A., & Boyer, D. (1999). Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and*

Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A), *21*(5), 701–708. doi: 10.1076/jcen.21.5.701.868

Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, *15*(5), 415-424.

Suhr, J., Hammers, D., Dobbinsbuckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology*, *23*(5), 521–530. doi: 10.1016/j.acn.2008.05.003

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, *14*(3), 189–207. doi: 10.1080/09084280701509083

Thames, A. D., Sayegh, P., Terashima, K., Foley, J. M., Cho, A., Arentoft, A., … Bookheimer, S. Y. (2016). Increased subcortical neural activity among HIV+ individuals during a lexical retrieval task. *Neurobiology of Disease*, *92*, 175–182. doi: 10.1016/j.nbd.2015.10.017

Tombaugh, T. N. (1996). *Test of Memory Malingering*. New York: Multi-Health Systems.

Trites, R. L. (1977). *Grooved pegboard.* Ottawa, Canada: Royal Ottawa Hospital.

Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, *27*(8), 1234–1246. doi: 10.1080/13854046.2013.835447

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*(2), 297–313. doi: 10.1080/13854040802232682

Whiteside, D., Clinton, C., Diamonti, C., Stroemel, J., White, C., Zimberoff, A., & Waters, D. (2010). Relationship between suboptimal cognitive effort and the clinical scales of the Personality Assessment Inventory. *The Clinical Neuropsychologist*, *24*(2), 315–325. doi: 10.1080/13854040903482822

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test (4th ed.)*. Lutz, FL: Psychological Assessment.

Willen, E. J., Cuadra, A., Arheart, K. L., Post, M. J. D., & Govind, V. (2017). Young adults perinatally infected with HIV perform more poorly on measures of executive functioning and motor speed than ethnically matched healthy controls. *AIDS Care*, *29*(3), 387–393. doi: 10.1080/09540121.2016.1234677

Wilson, T. W., Heinrichs-Graham, E., Robertson, K. R., Sandkovsky, U., O'Neill, J., Knott, N. L., … Swindells, S. (2013). Functional brain abnormalities during finger-tapping in HIV-infected older adults: A magnetoencephalography study. *Journal of Neuroimmune Pharmacology*, *8*(4), 965–974. doi: 10.1007/s11481-013-9477-1

Woods, S. (2004). Qualitative aspects of verbal fluency in HIV-associated dementia: a deficit in rule-guided lexical-semantic search processes? *Neuropsychologia*, *42*(6), 801–809. doi: 10.1016/j.neuropsychologia.2003.11.010

Woods, S., Scott, J., Dawson, M., Morgan, E., Carey, C., Heaton, R., & Grant, I. (2005). Construct validity of Hopkins Verbal Learning Test—Revised component process measures in an HIV-1 sample. *Archives of Clinical Neuropsychology*, *20*(8), 1061–1071. doi: 10.1016/j.acn.2005.06.007

Wu, J., Wu, H., Lu, C., Guo, L., & Li, P. (2015). Self-reported sleep disturbances in HIV-infected people: a meta-analysis of prevalence and moderators. *Sleep Medicine*, *16*(8), 901–907. doi: 10.1016/j.sleep.2015.03.027

Zhao, T., Wei, B., Long, J., Tang, X., Zhou, M., & Dang, C. (2015). Cognitive disorders in HIV-infected and AIDS patients in Guangxi, China. *Journal of NeuroVirology*, *21*(1), 32–42. doi: 10.1007/s13365-014-0295-x

APPENDIX A: INTAKE QUESTIONNAIRE

Gender: Female ☐ Male ☐ Other ☐

Age: _____

Handedness: Right ☐ Left ☐ Ambidextrous ☐ (i.e., able to use both hands with equal ease)

Years of Education: _____

1. Have you ever been diagnosed with one of the following?

a) Neurological disorder (e.g. stroke, multiple sclerosis)?
Yes ☐ No ☐

b) Have you ever had a traumatic brain injury or concussion?
Yes ☐ No ☐

2. Are you currently experiencing severe anxiety, depression, manic symptoms?

Yes ☐ No ☐

3. Do you have a history of trauma?

Yes ☐ No ☐

If yes, physical ☐ emotional ☐ sexual ☐ prefer not to say☐

APPENDIX B.1: INSTRUCTIONS FOR SIMULATED MALINGERING GROUP

Imagine that you were in a car accident in which another driver hit your car. You were knocked unconscious, and woke up in the hospital. The doctors told you that you had some bleeding in your brain after the accident.

Because the other driver is at fault, you have decided to take legal action against the driver. Your lawyer said that you may get more money if you look like you have sustained significant injuries because of the accident. You have decided to fake or exaggerate symptoms of a brain injury in order to increase the settlement you will receive. You have been told that common symptoms after a brain injury include difficulties with memory, concentrating, and being slower in responding.

The other driver's lawyer requires you to complete cognitive testing to determine if you sustained significant symptoms because the car accident. You know you can win a better settlement if you can convince the examiner that you have experienced significant brain damage. But if the examiner detects that you are faking, you are likely to lose the lawsuit.

You are about to take a series of cognitive tests that would be used in such a situation. I would like you to pretend you have brain damage, but in a believable way, such that your examiner cannot tell that you are attempting to fake a brain injury.

APPENDIX B.2: INSTRUCTIONS FOR NON-MALINGERING GROUP

You are about to take a series of cognitive tests. Some of the tests are easy and some are hard. I would like you to try your best on all of the tests.

APPENDIX C.1: POST-SESSION QUESTIONNAIRE FOR NON-MALINGERING CONTROL CONDITION

Discuss briefly what you were asked to do in this study:

_____

_____

How much did you try to follow the instructions during testing?

0 ---------1---------2---------3---------4---------5---------6---------7---------8---------9---------10

Did not try at all

Tried my absolute best

APPENDIX C.2 POST-SESSION QUESTIONNAIRE FOR SIMULATED MALINGERING CONDITION

Discuss briefly what you were asked to do in this study:

_____

_____

How much did you try to follow the instructions during testing?

0 ---------1---------2---------3---------4---------5---------6---------7---------8---------9---------10

Did not try at all

Tried my absolute best

How much could you imagine or relate to the motor vehicle accident scenario described?

0 ---------1---------2---------3---------4---------5---------6---------7---------8---------9---------10

Not at all

I could imagine it very vividly

What did you do during testing to pretend that you had cognitive difficulties? (circle as many as applies)

A. I responded to questions and completed tasks slower than usual
B. I answered questions incorrectly even though I knew the answer
C. I acted confused on how to complete the task
D. I asked the examiner to repeat questions
E. I didn't follow the test instructions
F. I didn't pretend
G. Other: _____

VITA AUCTORIS

Jessica Hurtubise was born in 1991 in Calgary, Alberta. She attended the University of Saskatchewan and completed a double major in Physiology and Pharmacology (B.Sc.) and Psychology (B.Sc.). She obtained a M.Sc in Physiology from the University of Saskatchewan in 2016. She is currently working toward obtaining a Doctoral degree in Clinical Psychology at the University of Windsor with hopes to graduate in Fall 2022.