# Investigation of Performance and Symptom Validity Testing in Children Utilizing Control, Simulation, and Clinical Groups

Elisabeth Vogt
*Marquette University*

# INVESTIGATION OF PERFORMANCE AND SYMPTOM VALIDITY TESTING IN CHILDREN UTILIZING CONTROL, SIMULATION, AND CLINICAL GROUPS

by

Elisabeth M. Vogt

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

September 2018

**ABSTRACT**

Investigation of Performance and Symptom Validity Testing in Children Utilizing Control, Simulation, and Clinical Groups

Elisabeth M. Vogt

Marquette University 2018

Integral to neuropsychology, assessment relies on valid self-report and credible performance on neuropsychological tests. Symptom exaggeration and misrepresentation of abilities confound interpretation of neuropsychological test data, subsequent diagnosis, and treatment. Measures evaluating performance and symptom validity have been extensively studied in adult populations; however, similar research in child and adolescent populations is limited. In accordance with recommended research methodology, this study utilized a simulation design with community recruited and medical center clinical criterion groups, which included 191 children and adolescents (7 to 16 years old). Sensitivity, specificity, and proposed cut-off scores are described for the Victoria Symptom Validity Test, Digit Span Age Corrected Scaled Score, Reliable Digit Span, Reliable Digit Span-Revised, Rey Fifteen Item Test, and Automatized Sequences Task. Novel embedded performance validity indicators for WRAML-2 Verbal Learning were developed, and cut-off scores are proposed for recognition discriminability, d prime, and forced choice measures. Additionally, symptom validity scales from parent- and child-report questionnaires suggested that select parent-report BRIEF and BASC-2 and child-report BASC-2 validity scales distinguished simulators from control and clinical participants. This study meaningfully and substantially adds to the current understanding of objective validity measurement in youth neuropsychological assessment and provides a framework for future development and investigation of youth performance validity tests and youth and parent symptom validity tests.

# TABLE OF CONTENTS

**TABLE OF CONTENTS (continued)**

# LIST OF TABLES

# Investigation of Performance and Symptom Validity Testing in Children Utilizing Control, Simulation, and Clinical Groups

Neuropsychological assessment examines the brain-behavior relationship with specific focus on cognition, and emotional and behavioral functioning. Neuropsychological assessment has a rich history that evolved from the convergence of multiple fields, such as philosophy, science, and medicine, and continues to progress (Lezak, Howieson, Bigler, & Tranel, 2012). Today as an applied science, clinical neuropsychology focuses on the behavioral manifestation of cognitive impairment. Moreover, neuropsychological evaluation allows for the assessment of functioning across multiple cognitive domains that may be impaired due to neurological injury or disease (Bianchini, Mathias, & Greve, 2001). Neuropsychological evaluations inform clinicians and patients of a wide variety of important diagnostic and treatment-related issues that may include information about the presence or level of cognitive impairment, and the degree to which an individual is able to participate in treatment or other daily living activities (Schoenberg & Scott, 2011). This project specifically focused on neuropsychological assessment of children and issues that may affect the evaluation.

## Issues That may Affect Test Performance and Symptom Report

There is an appreciation that misrepresentation of ability and symptom exaggeration are important issues to consider during test interpretation (Heilbronner et al., 2009). Misrepresentation of ability refers to the examinee's inaccurate, diminished performance on neuropsychological tests. Symptom exaggeration refers to the act in which an examinee (or someone reporting on behalf of the examinee) over reports, or

fabricates, symptoms. Accurate reporting and credible performance are essential to accurately interpret neuropsychological assessment results. Non-credible performance or exaggeration of cognitive impairment results in an inaccurate quantitative representation of an individual's actual abilities. Notably, performance and symptom invalidity are far more predictive of impairment on neuropsychological testing than severity of traumatic brain injury documented by neuroimaging and observed behavioral functioning (Lange, Pancholi, Bhagwat, Anderson-Barnes, & French, 2012). Thus, if not detected, performance and/or symptom invalidity confounds the diagnostic process and subsequent treatment for child, adolescent, and adult populations. In fact, inappropriate treatment may exacerbate an individual's symptoms (Kirkwood, 2012) and result in inappropriate use of limited healthcare, educational, and other societal resources (Horner, VanKirk, Dismuke, Turner, & Muzzy, 2014).

Neuropsychological evaluations are especially vulnerable to response distortion because the process requires cooperation and accurate reporting from the patient (Bianchini et al., 2001). To highlight the need for evaluation of validity in neuropsychological assessments, consensus statements by prominent neuropsychological organizations were developed to summarize empirical literature and to provide assessment, diagnostic, and general practice guidelines for validity assessment (American Academy of Clinical Neuropsychologists (AACN) Board of Directors, 2007; Bush et al., 2005; Heilbronner et al., 2009). Further, recent case series have encouraged the incorporation of performance and symptom validity testing into child and adolescent assessments (McCaffrey & Lynch, 2017). Given that appropriately validated tools and tests must exist to provide quantifiable evidence of valid performance and symptom

report, clinical research related to validity assessment has drastically increased in recent years in adult populations (Berry & Nelson, 2010) and in child and adolescent populations (DeRight & Carone, 2015; Kirkwood, 2015).

## Development of Deception Abilities

The development of the ability to misrepresent oneself or engage in deception is fundamentally relevant to research in child and adolescent performance and symptom validity testing. Development of deceptive abilities occurs throughout childhood. Some have erroneously concluded that children are not able to alter their performance or symptom report in a manner that would be considered valid (DeRight & Carone, 2015). In contrast to this opinion, developmental research suggests that abilities related to deception begin to develop in toddlerhood, and the ability to deceive significantly improves through early childhood. By around age eight, children can deny transgressions and consciously create false beliefs in others (Talwar & Lee, 2002). They are also able to sustain a lie by inhibiting verbal or non-verbal disclosure of deception and evade entrapment questions (Talwar, Gordon, & Lee, 2007). After age eight up to adolescence, deception becomes more sophisticated related to the development of executive functioning abilities (Anderson, 2002; Walczyk, Roper, Seemann, & Humphrey, 2003). By adolescence, deception skills are hypothesized to be similar to adults (Salekin, Kubak, & Lee, 2008). Thus, literature on the development of deception would suggest that by around age eight; children may be able to engage in response distortion.

Further, experimental research documents the ability of children and adolescents to alter their symptom report and performance under the direction of researchers on

neuropsychological testing in a manner that is believable (Baer, Kroll, Rinaldo, & Ballenger, 1999; Blaskewitz, Merten, & Kathmann, 2008; Faust, Hart, & Guilmette, 1988; Gunn, Batchelor, & Jones, 2010; McKinzey, Prieler, & Raven, 2003; Nagle, Everhart, Durham, McCammon, & Walker, 2006; Rambo, Callahan, Hogan, Hullman, & Wrape, 2015; Rogers, Hinds, & Sewell, 1996; Stein, Graham, & Williams, 1995). In a seminal study, Faust and colleagues (1988) documented the ability of children to simulate believable impairment on neuropsychological testing. The simulated profiles and actual clinical profiles obtained from patients were then sent to numerous neuropsychologists for evaluation of response distortion. None of the neuropsychologists judged the profiles to be abnormal due to response distortion. Or said another way, no neuropsychologist viewed the results as invalid. Thus, given that children in research settings can produce believable, feigned profiles, it is imperative that clinicians consider non-credible report or performance, and research be conducted to identify it.

## Contexts in Which Children and Their Parents Misrepresent Themselves

Multiple clinical cases of misrepresentation involving children also provide evidence that invalid symptom report or test performance occurs for numerous reasons (e.g. maintaining a sick role, avoiding legal consequences, securing monetary settlements). Those reasons may or may not be readily apparent to the clinician (Kirkwood, Yeates, Randolph, & Kirk, 2012). Invalid presentations are often identified in clinical contexts and research through diagnostic categories, namely conversion disorder, factitious disorder, and malingering (American Psychological Association; APA, 2013). For example, Libow (2000) identified 42 cases in which children and adolescents aged 8

to 18 years intentionally feigned (e.g., malingered) medical symptoms in order to assume sick roles for attention or to avoid other responsibilities (e.g. school attendance). Cases of malingering and malingering by proxy (or falsification of symptoms under the direction of someone else) also document the ability of children and adolescents to feign or exaggerate cognitive symptoms in neuropsychological evaluations in cases where external incentives are readily apparent. These clinical cases are documented in evaluations for various clinical conditions (e.g., Attention-Deficit/Hyperactivity Disorder; ADHD and mild Traumatic Brain Injury; mTBI; see Conti, 2004; Flaro, Green, & Blaskewitz, 2007; Kirkwood, Kirk, Blaha, & Wilson, 2010; Lu & Boone, 2002; Stutts, Hickey, & Kasdan, 2003), in disability evaluations (see Chafetz & Dufrene, 2014; Chafetz, & Prentkowski, 2011), and in forensic evaluations (see Flaro & Boone, 2009; Flaro et al., 2007).

Invalid performance or symptom report is not invariably associated with conversion disorder, factious disorder, or malingering categories. For example, Flaro and colleagues (2007) reported that a nine-year-old boy had much larger than expected performance discrepancies between two cognitive evaluations. The initial evaluation documented low average intelligence and the presence of a reading disability, whereas evaluation a year later documented superior intelligence and reading abilities. As an explanation for these discrepant findings, the boy reported that the previous examiner was “mean,” so he was mad and did not do his best on testing. In this case, objective measurements of performance validity may have prevented misdiagnosis, utilization of unneeded special education services, setbacks to the child’s education, and the child’s frustrated response to being pulled out of class for services. Knowledge of child and

adolescent performance and symptom validity as well as parental symptom validity could provide objective evidence of credibility in evaluations like those clinical examples just presented. Ultimately, this would improve clinical practice because more accurate diagnoses and treatment recommendations would be provided by clinicians.

Evidence related to base rates of invalid performance or symptom report in clinical, forensic or psychoeducational contexts has started to emerge. Notably, children and adolescents undergoing Social Security Administration disability evaluations for benefits display the highest rates invalidity; estimated at 26 to 60% of cases (Chafetz, 2008; Chafetz, Abrahams, & Kohlmaier, 2007). It appears that parents may meaningfully influence the invalid presentation in those social security evaluations. In studies with children and adolescents who sustained a mild Traumatic Brain Injury (mTBI), base rates of invalid performance range from 12 to 20% (see Araujo et al., 2014; Baker, Connery, Kirk, & Kirkwood, 2014; Green, Kirk, Connery, Baker, & Kirkwood, 2014; Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014; Kirkwood, Connery, Kirk, & Baker, 2014; Kirkwood, Hargrave, & Kirk, 2011; Kirkwood & Kirk, 2010; Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014; Kirkwood, et al., 2012). In mixed clinical populations, comprised of a variety of developmental, medical, and neurological conditions, base rates of invalid performance are typically estimated between 2 to 5% (see Donders, 2005; Green, Flaro, Brockhaus, & Montijo, 2012; Kirk, Harris, Hutaff-Lee, Koelemay, Dinkins, & Kirkwood, 2011; Ploetz, Mazur-Mosiewicz, Kirkwood, Sherman, & Brooks, 2014). Within college-aged samples undergoing psychoeducation evaluations, base rates of invalidity in ADHD evaluations are estimated at 15 to 47% (see Harrison & Edwards, 2010; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008; Sullivan,

May, & Galbally, 2007), whereas in learning disability evaluations, performance invalidity base rates are approximately 15% (see Harrison, Edwards, & Parker, 2008; Sullivan et al., 2007). Unfortunately, literature related to base rates of child and adolescent invalid performance in psychoeducational evaluations is not available, though multiple papers documenting case studies of non-credible presentations exist (see Harrison, Green, & Flaro, 2012; Kirkwood et al., 2010; Lu & Boone, 2002). Thus, across settings in which psychological evaluations occur, documentation of non-credible presentations occurs.

## Performance and Symptom Validity Paradigms

Larrabee (2012) proposed and defined the terms *performance* and *symptom validity* to provide clarity in validity research. These terms distinguish between self-report and performance test methods. Performance validity is the credibility of performance on a measure assessing cognitive ability. Symptom validity refers to the accuracy or truthfulness of symptom reporting on self-report measures. Actual corresponding tests are therefore performance validity tests (PVTs) and symptom validity tests (SVTs; Larrabee, 2012).

The overall goal of a PVT is to provide detection strategies for invalid cognitive performance. This detection involves non-credible, excessive impairment or an unlikely presentation that would be inconsistent with neuropsychological sequelae (Berry & Schipper, 2008). PVTs include stand-alone or embedded validity tests. A stand-alone PVT is a separate test specifically designed to assess credibility of performance, whereas, an embedded PVT reflects a psychometrically defined extreme cut score (e.g. a

benchmark that denotes passing or failure of a SVT or PVT) within a traditional neuropsychological test (Strauss, Sherman, & Spreen, 2006). Stand-alone PVTs are specifically developed to assess only the credibility of performance and designed to be exceptionally easy so that invalid performance is detected if test results are below a specific score or score range. As a result, diagnostic classifications may be optimized when using stand-alone measures (Bianchini et al., 2001), though the stand-alone measure must exhibit face validity consistent with actual measures of cognitive ability. The challenge in developing and utilizing embedded validity measures is that individuals with bona-fide impairment, or children and adolescents with developing cognitive abilities, may perform poorly on these measures due to actual relatively limited capabilities (Strauss et al., 2006). Subsequently, the floor must be low enough to avoid misidentifying an individual as providing invalid performance when his/her performance was credible (i.e., false positive classification). Thus, sensitivity of embedded PVTs may be moderate at best.

SVTs are frequently scales developed within measures of psychological symptoms and assess the validity of responses (e.g. inconsistent responding to similar items or endorsement of highly atypical symptoms; Strauss et al., 2006). However, some SVTs were developed as stand-alone measures to only quantify validity of self-report (e.g., Structured Interview of Reported Symptoms (SIRS), Second Edition for adults; Rogers, Sewell, & Gillard, 2010). Typical categories of test taking attitudes are identified using the following methods: a) consistency in responding to similar items, b) attempts to appear desirable by responding in a positive light to infrequently endorsed items, c)

efforts to create an overly negative, faked, or implausibly severe profile, and d) general engagement in responding (Hoelzle, Nelson, & Arbisi, 2012; Meehl & Hathaway, 1946).

## Research Methods Relevant to Performance and Symptom Validity Testing

Investigation of performance and symptom validity testing poses unique challenges. A key challenge is that it is impossible to recruit participants that are genuinely and unmistakably providing an invalid profile in clinical contexts (Bigler, 2012; Rogers, Harrell, & Liff, 1993). Individuals engaging in response distortion during clinical evaluations rarely, if ever, acknowledge doing so. As a result, PVT and SVT standardization studies must rely on simulation volunteers or other known clinical or forensic groups that exhibit a high likelihood of demonstrating invalid performance or symptom report to evaluate the clinical utility of measures (Brennan & Gouvier, 2006; Rogers, 2008).

Performance and symptom validity research has primarily focused on adult populations. Only in very recent years have studies included child and adolescent samples (DeRight & Carone, 2015). The two most prominent performance and symptom validity design methods in adult research are known-groups comparison and simulation studies (Berry & Schipper, 2008; Rogers, 2008). However, in child and adolescent PVT and SVT research, descriptive studies are relatively common. The descriptive method involves administering a PVT to a clinical group of children and adolescents and then reporting the percentage of patients that "pass" (i.e., score above a previously determined cut score). The cut-off scores employed are frequently drawn from adult research. If a child or adolescent achieves a "passing" score based upon the adult cut-off score, it is

then suggested that the patient demonstrates valid performance or symptom report (e.g., Brooks, 2012). However, in these PVT and SVT descriptive studies, the researcher is unable to classify groups as providing valid or invalid performance based upon an established child and adolescent criterion. Thus, it is unknown if validity test failures are due to invalid performance or to other developmental factors.

Conversely, the known-groups comparison design is a recommended method that includes individuals in real-world conditions that are classified by independent researchers according to their specific response style. This design may be utilized to determine cut-off scores and utility estimates of new PVTs and SVTs and to determine how performance and symptom invalidity are related to other factors (e.g. general neuropsychological test performance or daily functioning). For example, individuals in a clinical context with below-chance performance on two PVTs may be assigned to an invalid group, whereas individuals passing all PVTs would be assigned to the clinical control group (Slick, Sherman, & Iverson, 1999). A known-groups comparison approach includes two phases. First, groups are independently established based on whether the patient has provided a valid or invalid report or performance by, ideally, a well-established criterion. Then the research team, blind to group assignment, investigates similarities or dissimilarities in order to determine how well a new validity measure may function, or to determine the effect of performance or symptom validity on other factors (e.g. performance on cognitive tests; Rogers, 2008). External or ecological validity is strong because participants have tangible incentives and consequences associated with the outcome of the assessment (Berry & Schipper, 2008). However, the initial classification of participants is challenging because researchers must utilize multiple,

empirically supported measures of validity classification (e.g., evidence from multiple PVTs, SVTs, and record review), and consensus is determined about cut-scores or other indicators for accurate classification. Thus, well-established criterion for validity is necessary in order to determine the initial group assignment. This requirement poses a particular challenge within child and adolescent populations due to the limited amount of information currently available for diagnostic classification statistics for the majority of PVTs and SVTs (DeRight & Carone, 2015; Kirkwood, 2012; 2015). Thus far, published PVT cut-scores from adult samples or youth mTBI samples have been primarily utilized to differentiate groups in multiple child and adolescent samples (see Appendix A).

In response to limitations of descriptive and known-groups methods, simulation designs are recommended to initially investigate PVT and SVT diagnostic classifications prior to their use in known-groups comparisons (Rogers, 2008). The simulation design is an experimental method in which non-clinical participants are randomly assigned to a group a-priori, and those groups differ through instructions about how to complete study measures. The control group is asked to perform all tasks under standard procedures (e.g., to the best of one's ability), while the simulation group is given instructions to produce a specific response style consistent with a specific disorder or cognitive impairment (Rogers et al., 1993).

At the most basic level, simulation design includes only those two groups, simulators and controls. Thus, a common criticism of simulation research involves the assertion that simulators may not perform similarly to genuine patients who are feigning in clinical or forensic contexts since research participants do not encounter the same consequences associated with succeeding or failing PVTs and SVTs (Larrabee, 2012).

Consequently, external validity is weak, and it is unclear whether results of simulation research using only control and simulator groups can be generalized to clinical or forensic populations. On the other hand, the methodology of simulation designs is more complex than the designs previously described (e.g. descriptive studies and known-groups comparison). Given the need for a standardized protocol, numerous decisions for experimental procedures must be made in simulation research in regards to instructions, incentives, and manipulation checks, which can all directly affect validity, but provide opportunity for strong internal validity (Brennan & Gouvier, 2006; Rogers, 2008). In addition, causal inferences can be made when differences arise between experimental groups since results are effectively testable (Berry & Schipper, 2008).

Simulation research with only control and simulation groups has been effectively applied a number of times in child and adolescent assessment research. Utilizing non-clinical samples, five PVT and two SVT studies sought to validate measures with variations on the specific simulation design that included: (a) community recruited control and simulation groups (Blaskewitz et al., 2008; Gunn et al., 2010; Rambo et al., 2015) or (b) one non-clinical group of individuals that participated in both the control and simulation conditions (McKinzey et al., 2003; Nagle et al., 2006; Rogers et al., 1996; Stein et al., 1995). Developmental literature suggests that children are capable of distortion or feigning, though documentation of ability to feign cognitive impairment or exaggerate symptoms in a research context is critical to simulation methodology. In most studies, children and adolescents in the simulation condition were able to feign cognitive impairment (see Blaskewitz et al., 2008; Gunn et al., 2010; Lucio, Durán, Graham, & Ben-Porath, 2002; McKinzey et al., 2003; Rambo et al., 2015) or exaggerate symptoms

(see Baer et al., 1999). However, in one study in which 6- to 12-year-olds completed both the simulation and control conditions, children were not able to suppress performance on a PVT (Nagle et al., 2006). It appears that requesting children to engage in two conditions in one session may negatively affect the experimental manipulation. Thus, internal validity may be optimized when children and adolescents participate in either control or simulation conditions.

While the fore-mentioned studies documented child and adolescent ability to understand and comply with simulation conditions, it is unknown how children with clinical conditions (e.g. intellectual impairment) would perform on these same PVT or SVT measures. Documentation of PVT or SVT performance in children with clinical conditions is critical since performance should not be correlated with actual abilities (Strauss et al., 2006). Thus, exclusion of a clinical criterion group for comparison confounds results because it is unknown if individuals with bona-fide impairment would perform at a level consistent with simulators. In addition, it is unknown if the simulator group performance would be overly discrepant from a group with secondary gain incentives (e.g., forensic, compensation, or academic accommodations). Thus, while simulation design provides valuable information about non-clinical child and adolescent performance on PVTs and offers documentation of child and adolescent ability to feign, questions remain as to generalizability for clinical samples in which children and adolescents possess actual cognitive impairment or psychological disorders.

The optimal and most rigorous simulation design for investigating PVTs and SVTs includes control, simulation, *and* criterion groups. To date, two symptom validity studies investigating the Minnesota Multiphasic Personality Inventory – Adolescent

(MMPI-A; Butcher et al., 1992) utilized this method without engaging participants in more than one condition (Baer, et al., 1999; Lucio et al., 2002). No PVT simulation studies, to date, have included a criterion group. In Baer et al. (1999) and Lucio et al. (2002), validity scales effectively discriminated between adolescents in clinical, non-clinical control, and simulator groups providing confidence that the symptom validity scales could distinguish bona-fide patients from those providing invalid profiles. In addition, diagnostic classification statistics were derived for each validity scale at various cut-off scores, which made it possible to determine which scales and cut-off scores optimally differentiated between feigning and clinical groups. Simulation designs without criterion groups cannot effectively address how simulation results might compare to results from individuals in clinical settings who are engaging in some type of response distortion and who possess actual incentives or consequences; however, the fore-mentioned studies (Baer et al., 1999; Lucio et al., 2002) could determine generalization of cut-off scores. While challenges would exist for recruitment of a criterion sample, data from these forensic or clinical populations is critical for application of simulation study results.

In summary, multiple research designs are possible within PVT and SVT research (e.g. descriptive, known-groups comparison, and simulation). A primary research challenge is the inability to recruit individuals engaging in feigning due to real-world experiences and consequences (Bigler, 2012). Researchers are also faced with the challenge of balancing clinical relevance and experimental control, which can be inversely related (Rogers et al., 1993). Known criterion groups demonstrate better clinical relevance due to real-world situations that affect performance and symptom validity (e.g.,

external validity), whereas, simulation and control groups allow the researcher to maximize internal validity, but participants do not experience the meaningful consequences or benefits associated with a clinical or forensic evaluation. Thus, the gold-standard approach to evaluate PVTs and SVTs involves use of a control, simulation, and a criterion group (e.g. either forensic or clinical) to address internal and external validity concerns (Heilbronner et al., 2009; Rogers, 2008). Then, once criterion for valid report or test performance is reliably established, known-groups comparison designs are recommended (Larrabee, 2012). A *very* limited number of child and adolescent PVT and SVT studies are available that reflect recommended methodological approaches, thus, additional research is greatly needed. This study aims to fill that methodological gap in child and adolescent PVT and SVT literature.

## Aims

To date, limited literature is available describing empirically derived cut-off scores for PVTs and SVTs in child and adolescent samples, despite documentation that children can engage in feigning in research settings and case studies demonstrating suppression of test performance or misrepresentation of symptoms to attain external rewards in clinical settings. Multiple PVT research studies are descriptive studies in which cut-off scores established with adults are applied to child and adolescent populations. For multiple measures, children can perform at a level consistent with valid performance in adults; however, given the methodological design of these studies, it is unclear if failures are due to invalid performance or related to other developmental factors. Further, children under 10 years are more likely to be identified as providing non-

credible performance on some PVTs, so it seems that cut-off scores or measures may not be appropriate for all age groups (Constantinou & McCaffrey, 2003; Courtney, Dinkins, Allen, & Kuroski, 2003; Donders, 2005; Perna and Loughan, 2013; Ploetz et al., 2014).

Therefore, **Aim 1** of the present study is to report PVT and SVT performance of children from clinical and non-clinical groups (e.g. control, simulation, community-clinical, and medical center-clinical). Specifically, descriptive statistics, associations with age and intellectual functioning, and the percentage of children that were able to "pass" PVTs at adult or available child/adolescent cut-off scores is reported. It was hypothesized that the majority of children in the control and clinical groups would "pass" probabilistic measures of performance validity, whereas, younger children (e.g. under 10 years old) and children with greater degrees of cognitive impairment would demonstrate greater rates of PVT failure on embedded measures when using adult cut-off scores.

Average performance of children in the simulation group was compared to the control, community-clinical, and medical center-clinical groups to determine if those children in the simulation group were able to suppress their performance on neuropsychological tasks and alter their responses on questionnaires to present with impairments (i.e., perform at a level suggesting "clinical" impairment). Based upon previous literature, it was hypothesized that children in the simulation group would be able to engage in response distortion and their PVT and SVT performance could be differentiated from control and clinical participants.

Utility estimates (e.g. sensitivity and specificity) are exceptionally limited for the majority of PVTs and SVTs investigated, to date, in children. The strength of the simulation study approach is that it provides the opportunity for specific benchmarks of

validity to be established and their effectiveness for differentiation of key groups to be known. Subsequently, **Aim 2** involved identification optimal cut scores for each included PVT and SVT utilizing the control, simulation, community-clinical, and medical center-clinical groups. Based upon previously described literature, it was expected that stand-alone PVTs would demonstrate better sensitivity and specificity than embedded measures.

Additionally within Aim 2, multiple PVT measures were compared to determine which measures demonstrated optimal diagnostic classification cut-off scores. Previous literature suggested that the stand-alone measure the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997) would demonstrate better utility estimates than embedded measures; thus it was hypothesized that the VSVT would exhibit the strongest sensitivity and specificity when comparing stand-alone and embedded PVTs.

To date, no PVT and SVT study with children has included parental report, nor has pediatric symptom validity been investigated in conjunction with performance validity in simulation research. VanDyke and colleagues (2013) reported that invalidity on a PVT does not necessarily equate with invalidity on symptom validity scales. While PVTs and SVTs are complimentary and help clinicians make decisions regarding the validity of a patient's presentation, they are believed to assess different constructs. Thus, **Aim 3** was to investigate the relationship between PVT and SVT validity in children in order to evaluate which measures converge and if certain tests possess greater utility. It was hypothesized that PVTs, relative to SVTs, would more effectively differentiate the control, simulation, community-clinical, and medical center-clinical groups. Further, it

was hypothesized that the parent/guardian would be able to simulate, or alter their report of child/adolescent emotional and behavioral functioning according to research instructions, since deception skills should be developed by adulthood. However, since no previous literature has investigated simulated parental report, it is unknown if SVT scales on parent-report measures will be useful for differentiation of control, clinical, and simulation parent groups.

## Method

### Participants

Four groups were included in this study. Three groups were recruited from the community. Children without a previously identified psychological diagnosis (e.g., ADHD, autism spectrum disorder, or intellectual disability) were assigned to the control or simulation groups. Whereas, children recruited from the community that had a previous diagnosis were within the community-clinical group. Inclusion criteria for the community-clinical group involved a previous DSM-5 diagnosis or a neurological condition (e.g., epilepsy) in the absence of a DSM-5 diagnosis. The fourth group included children seen for a clinical evaluation due to concerns for cognition in the context of neurological, medical, or developmental conditions. In order to identify participants for the fourth group, a retrospective chart review was completed to identify children evaluated in an academic medical center pediatric neuropsychology department that completed the VSVT as a part of their standard evaluation. The primary goal of the inclusion of two clinical groups (e.g. community-clinical and medical center-clinical) was

to ensure that criterion groups with varying levels of likely cognitive impairment were available for comparison and that one criterion group (e.g. community-clinical) did not possess any identifiable benefits from an evaluation.

**Community Recruitment.**

For community recruitment, 8 to 16-year-old children and adolescents and their parent/legal guardian were recruited from community schools and through general advertisements. Institutional review board approval was obtained from Marquette University and the Medical College of Wisconsin. The principal investigator screened participants prior to scheduling a testing session. Importantly, parents and guardians were informed that participation in the research session would not constitute a diagnostic or academic accommodation evaluation and that cognitive test scores would not be provided in a report. This criterion, that the research session would not also constitute a clinical evaluation, decreased or eliminated the possibility of secondary gain issues impacting test performance or questionnaire responses. Children that had a prior psychological or neurological diagnosis (e.g. mood, behavior, or neurodevelopmental disorders) were assigned to the community-clinical group. Inclusion of this community-clinical criterion group ensured that the internal validity of measures administered was controlled so that direct comparisons could be made with the control and simulation groups. Children recruited from the community without a previous diagnosis were assigned to either the control or simulation group stratified by sex and age.

Test sessions for community-recruited participants were completed at the Marquette University Center for Psychological Services. Since emotional and behavioral questionnaires were completed by the parent/guardian and the child or adolescent, critical

items from the Behavior Assessment System for Children-Second Edition (BASC-2; Reynolds & Kamphaus, 2004) were reviewed. When critical items were endorsed, the graduate student investigator conducted a risk assessment, discussed the item response with the parent/guardian, and followed Marquette University Center for Psychological Services procedures for self and/or other harm. During the course of the study, four children endorsed critical items. On two occasions, the child did not understand that the question referred to self-harm. On the other two occasions, the parent and child both reported awareness of self-harm desire during anger or frustration, and no evidence of suicidal intent were identified. Community referrals for therapy were provided to the parent/legal guardian in those cases and upon request from other families. After the test session was completed, the child selected a prize and received a Junior Researcher Certificate and a gift certificate to a restaurant. The parent/legal guardian received $20 cash.

**Standard instructions.**

Parents and children in the community-clinical and control groups completed test measures and questionnaires according to standard testing procedures. After consent and assent were attained, parents were instructed to complete their questionnaires according to the standard manual instructions. Children received the following instructions prior to the start of their test session, “You’ll be doing a lot of things today, like looking at pictures, answering questions, and completing a task on the computer. Some things may be easy for you, but some may be hard. Just try your best.” After the test session, dyads in the clinical and control groups were thanked for their participation.

**Simulation instructions.**

Following consent and assent procedures, children and adolescents in the simulation group completed three measures under standard conditions utilizing the same instructions that were given to the control and community-clinical groups. These measures were administered under standard conditions to ensure ability to identify numbers (e.g. the Bracken School Readiness Assessment; Bracken, 2007) and to obtain an estimate of intellectual functioning (e.g. Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) Vocabulary and Matrix Reasoning subtests; Wechsler, 2014). Meanwhile, the parent/guardian completed a demographics form in the lobby.

Following those three tasks, the parent/guardian was asked to return to the testing room. The child or adolescent and their parent/guardian were informed that they were selected to be in the experimental group with special instructions (see *Appendix B* for child/adolescent and *Appendix C* for parent/guardian verbatim instructions). Consistent with simulation methodology recommendations, instructions were created and tested for clarity, ease of understanding, and a relatable scenario that provided context (Bianchini et al., 2001; Rogers, 2008). The child or adolescent was read a scenario in which they were asked to feign cognitive impairment as a means to reduce their workload at school. Comprehension was evaluated with a practice trial. Children and adolescents were also asked to describe instructions to the examiner to ensure instructions were understood. The parent/guardian received a similar scenario and was asked to respond to questionnaires in a manner that would assist their child in receiving academic accommodations.

In accordance with simulation methodology suggestions, debriefing and a manipulation check were utilized to, respectively, provide context and limit non-credible performance to the study and evaluate the participants' recall, comprehension, and reported compliance with instructions (Bianchini et al., 2001). Following the test session, both the parent/guardian and child in the simulation group were informed of the study purpose and rationale for simulation (see *Appendix D* for verbatim debrief). Additional questions were addressed, as needed, regarding the experimental condition and aims of the research study. The child/adolescent and parent/guardian also completed a brief manipulation check, in which they rated their understanding of instructions and how accurately they followed the simulation instructions (see *Appendix E*).

**Academic Medical Center Recruitment.**

Recruitment of the medical center-clinical group entailed a retrospective chart review of patients that were seen for a clinical neuropsychological evaluation between January 1, 2016 through March 31, 2017 in the Medical College of Wisconsin (MCW) Pediatric Neuropsychology Department. These patients were referred for varied neurological concerns (e.g., seizure disorders), as a part of developmental follow-up clinics (e.g., cardiac conditions), or for general cognitive and behavioral concerns (e.g., attention problems). The purpose of the evaluations was to determine if the patient met criteria for a diagnosis and to provide treatment recommendations. The VSVT had been obtained for the department and providers could include the measure in the evaluation. Patients eligible for inclusion involved all children and adolescents, 7 to 16 years of age, that were administered the VSVT during their evaluation. Specific measures (Rey FIT,

AST, WRAML-2 Forced Choice Measure) included in the community-recruited sample were not administered to patients in the medical-center sample given that it would lengthen the evaluation. Further, parent questionnaires included in the community-recruited sample were not routinely utilized in the medical-center group. The primary measure of interest was the VSVT. The Children's Hospital of Wisconsin Institutional Review Board approved the retrospective chart review.

**Sample.**

Child and adolescent participants [*N* = 191, *M*age = 12.19 (2.80), age range 7.10 to 16.11 years] in this study included community-recruited child/adolescent and parent/guardian dyads (*n* = 66) and medical center patients seen for a neuropsychological evaluation (*n* = 125). Previous research indicates that very large effect sizes (e.g., Cohen's *d* > 1.00) are consistently observed in adult simulation research (Rogers, 2008) to validate PVTs and SVTs with control, clinical, and simulation groups. A-priori power analysis utilizing G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) and the smallest observed effect size (Cohen's *d* = 1.23, see Blaskewitz et al., 2008; Gunn et al., 2010; Nagle et al., 2006; Rambo et al., 2015) in child and adolescent simulation research indicated a sample size of 12 was needed to detect group differences in PVT and SVT performance for a two-tailed *t*-test with power set at 0.80, though a more conservative effect size (Cohen's *d* = 1.00) indicated a sample of 17 participants was needed in each group. Post-hoc power analysis varied across analyses, however, for multiple PVTs (e.g., VSVT, Reliable Digit Span) observed power in the present study was generally > .92 to

detect a large effect size (Cohen's $d$ = 1.00), which indicated the present sample size was sufficient for research aims.

Community-recruited children and adolescents [$n$ = 66, *M*age = 10.55 (2.50); age range 8.00 to 16.30; 54.5% male] all spoke English as their first language. Parents/guardians predominantly spoke English as their first language (98.5%, $n$ = 65). Children/adolescents (Caucasian 69.7%, African American 18.2%, Bi/Multiracial 7.6%, Asian 1.5%, Native Hawaiian/Pacific Islander 1.5%, and Other 1.5%) and parent/guardians (Caucasian 71.2%, African American 18.2%, Bi/Multiracial 4.5%, Hispanic/Latino 1.5%, Asian 1.5%, and Other 3%) from the community primarily identified as Caucasian. They were subdivided into control [$n$ = 23, *M*age = 9.82 (2.10); 52.2% male], simulation [$n$ = 20, *M*age = 11.72 (2.62); 50% male], and clinical [$n$ = 23, *M*age = 10.25 (2.52); 60.9% male] groups. Per study procedures, parent/guardians were assigned to the same group as their child in order to receive standard (control $n$ = 23 and clinical $n$ = 23) or simulation ($n$ = 20) instructions. Children and adolescents were assigned to the community-clinical group if they had a pre-morbid diagnosis (see Table 1). Of note, over half of the community-recruited clinical group had more than one psychological diagnosis [one diagnosis $n$ = 10 (43.5%); two diagnoses $n$ = 6 (26.1%); three diagnoses $n$ = 3 (13%); four diagnoses $n$ = 3 (13%); and six diagnoses $n$ = 1 (4.3%)]. The majority of participants in the community-clinical group had an Individualized Education Plan (IEP) or 504 Plan (56.5%, $n$ = 13) and some participants (39.1%, $n$ = 9) had taken a prescribed psychotropic medication prior to the research session.

Table 1
*DSM-5 Diagnoses of Clinical Participants*

| | Community-Clinical (*n* = 23) | Medical Center -Clinical (*n* = 125) |
|---|---|---|
| ADHD | 52.2% (12) | 38.4% (48) |
| Autism | 21.7% (5) | 4% (5) |
| Intellectual Disability | 17.4% (4) | 6.4% (8) |
| Learning Disorder | 21.7% (5) | 12% (15) |
| Speech/Language | 39.1% (9) | 6.4% (8) |
| Depressive Disorder | 9.7% (2) | 20.8% (26) |
| Anxiety Disorder | 52.2% (12) | 19.2% (24) |
| Trauma/Stress Disorder | 0% | 8% (10) |
| Oppositional Defiant | 4.3% (1) | 0.8% (1) |

*Note:* DSM-5 = Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition, ADHD = Attention-Deficit/Hyperactivity Disorder

Retrospective chart review was conducted for children and adolescents [*N* = 125, *M*age = 13.06 (2.55), age range 7.10 to 16.11 years, 48.8% male] that completed a neuropsychological evaluation at MCW and had a complete VSVT. Children/adolescents within the medical center-clinical group were predominantly Caucasian (52%; African American 15.2%, Hispanic/Latino 8%, Bi/Multi-racial 8%, Asian 1.6%, Other 0.8%, and Race/ethnicity not available 14.4%). Nearly half (48.8%, *n* = 60) of participants from the medical center-clinical group had an IEP or 504 Plan prior to their neuropsychological evaluation, and a quarter (24.8%, *n* = 31) had a prescribed psychotropic medication. Psychological disorder diagnoses from DSM-5 are described in Table 1. Of note, a quarter of the medical-center clinical group had more than one psychological diagnosis [No diagnosis *n* = 25 (20%); one diagnosis *n* = 67 (53.6%); two diagnoses *n* = 21 (16.8%); three diagnoses *n* = 10 (8%); and four diagnoses *n* = 2 (1.6%)]. The majority of children and adolescents (*n* = 100) in the medical center-clinical group also had a medical or neurological diagnosis (Epilepsy 24%, *n* = 30; mild Traumatic Brain Injury 20.8%, *n* = 26; moderate Traumatic Brain Injury 2.4%, *n* = 3; Brain tumor or lesion 10.4%, *n* = 13;

Congenital Heart Defect 3.2%, *n* = 4; Genetic Disorder 3.2%, *n* = 4; Encephalitis 3.2%, *n* = 4; other neurological condition 12.8%, *n* = 16). Of note, neurological diagnoses that represented less than 2% of the medical center-clinical sample were combined into the 'other neurological' category and included, but were not limited to, multiple sclerosis, spina bifida, cerebral palsy, and hydrocephalus.

**Measures**

Community-recruited children and adolescents completed the Bracken Number Identification task, subtests to estimate intellectual ability (WISC-V Vocabulary and Matrix Reasoning), embedded and stand-alone PVTs [WISC-V Digit Span Reliable Digit Span (RDS); Greiffenstein, Baker, & Gola, 1994; Reliable Digit Span-Revised (RDS-R); Spencer, Tree, Drag, Pangilinan, & Bieliauskas, 2010; Automatized Sequences Task (AST); Kirkwood et al., 2014; Rey Fifteen-Item Test (Rey FIT); Rey, 1964; Wide Range Assessment of Memory and Learning-Second Edition (WRAML-2) Verbal Learning subtest; Sheslow & Adams, 2003; VSVT)] and a mood and behavior questionnaire that includes symptom validity scales (BASC-2). To avoid test order effects, four different test orders were created and assigned randomly, though the WISC-V Vocabulary and Matrix Reasoning subtests and the Bracken Number Identification subtest were always administered first due to the simulation condition (see Appendix F for test orders). The parent/guardian completed two parent-observation mood and behavior questionnaires that contain symptom validity scales [BASC-2; Behavior Rating Inventory of Executive Function (BRIEF); Gioia, Isquith, Guy, & Kenworthy, 2000].

Children and adolescents from the medical center-clinical group completed the VSVT, and the majority completed the WISC-V Vocabulary (*n* = 120, 96%), Matrix Reasoning (*n* = 117, 94%), and Digit Span (*n* = 122, 97%) subtests. The majority of participants were administered the complete WISC-V, though some had completed an abbreviated WISC-V or another intelligence measure. A portion of the medical center-clinical group (*n* =54, 43%) also completed the WRAML-2 Verbal Learning subtest.

***Bracken School Readiness Assessment.*** The Bracken School Readiness Assessment – Third Edition (Bracken, 2007) is utilized to determine a child's understanding of early academic concepts related to color, letter, shape, and number identification. While the normative data is not relevant to the age group included in the proposed study (8 to 16 years of age), the Number Identification subtest was utilized to ensure that children included in the study possessed basic number knowledge since this skill is needed to complete PVT tasks.

***Wechsler Intelligence Scale for Children – Fifth Edition***. Two subtests from the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V; Wechsler, 2014) were included to estimate verbal and non-verbal abilities. In the Vocabulary subtest, the child or adolescent was read a word and was asked to describe the meaning of the word. The vocabulary subtest assesses ability to access and express word knowledge. In the Matrix Reasoning task, the child was shown a pattern with a missing piece and selected a response from five options. Matrix Reasoning assesses non-verbal conceptual relationships utilizing inductive reasoning.

***Reliable Digit Span.*** Digit Span is a frequently administered WISC-V subtest that assesses verbal attention and working memory. Reliable Digit Span (RDS; Greiffenstein

et al., 1994) is an embedded effort measure in which a floor effect cut-off score was developed for evaluation of adult performance validity. RDS is calculated by summing the longest string of digits forward and backward in which there were no errors across two trials. Many individuals with well-documented cognitive impairment can repeat strings of digits consisting of 3, 4, or 5 digits reliably, thus failure to do so might suggest insufficient engagement. In adults, an RDS ≤ 7 indicates invalid performance. An additional cut score utilizing the achieved age-corrected scaled score (Digit Span Age-Corrected Scaled Score; DS ACSS ≤ 5) from Digit Span has also been developed in adult samples to provide an alternative validity measure (Axelrod, Fitchenberg, Millis, & Wertheimer, 2006). In a sample of children and adolescents with mTBI, Kirkwood and colleagues (2011) established new cut-off scores (RDS ≤ 6 and ACSS ≤ 5) that resulted in moderate sensitivity and strong specificity. In contrast, in a more heterogeneous clinical sample of children and adolescents (Perna, Loughan, Hertza, & Segraves, 2014) and in a sample with children and adolescents with dual diagnoses (Loughan, Perna & Hertza, 2012), an optimal RDS cut-off score was ≤ 4. However, the higher pediatric RDS cut-off is more commonly referenced (see Araujo et al., 2014; Welsh, Bender, Whitman, Vasserman, & MacAllister, 2012).

***Reliable Digit Span-Revised.*** The WISC-V added a sequencing trial to Digit Span, in contrast previous versions of the WISC that only included forward and backward trials. Reliable Digit Span-Revised (RDS-R; Spencer et al., 2010) adds the reliable span (the longest string of numbers correct across two trials in the same set) from sequencing to reliable spans from forward and backward trials. Investigations with adult samples have suggested a RDS-R cut-off score of ≤ 11 was optimal for differentiating

undergraduates in control and simulation conditions (Reese, Suhr, & Riddle, 2012), Veterans that sustained a mTBI with valid or invalid test performance (Spencer, Axelrod, Drag, Waldron-Perrine, Pangilinan, & Bieliauskas, 2013), and Veterans with mixed clinical conditions with valid or invalid test performance (Young, Sawyer, Roper, & Baughman, 2012).

***Automatized Sequences Task.*** The Automatized Sequences Task (AST) was derived from the Sequences Task in the Children's Memory Scale (CMS; Cohen, 1997), which assesses processing speed and the ability to mentally manipulate and sequence rote verbal information. The AST involves completing four basic tasks as quickly as possible (e.g., reciting the alphabet, days of the week, months of the year and counting to 20). Kirkwood and colleagues (2014) identified that saying the alphabet (≥ 8 seconds), counting to 20 (≥ 6 seconds), days of the week (≥ 4 seconds), months of the year (≥ 10 seconds), and total test time (≥ 27 seconds), generally exhibited moderate sensitivity to invalid performance in a mTBI sample of 8- to 17-year-olds.

***Rey Fifteen-Item Test.*** The Rey FIT (Rey, 1964) is a stand-alone validity test that requires reproduction of familiar and repetitive stimuli (e.g. A, B, C), and thus, it is designed to be very simplistic so that even individuals with cognitive impairment can reproduce nearly all of the stimuli on the recall task. The number of items correctly recalled from the stimulus card is interpreted for validity, which in adult normative studies has resulted in divergent recommendations (Rey FIT total score ≤ 7, ≤ 8, ≤ 9, ≤ 11) with higher benchmarks resulting in better sensitivity but lower specificity (Strauss et al., 2006). Child and adolescent literature suggested cut-off scores of ≤ 7 for non-clinical 6- to 11-year-olds and < 9 for 8-to 17-year-olds with mTBI (Green et al., 2014).

However, younger children (< 10 years) exhibit generally poorer performance (Constantinou & McCaffrey, 2003)

Boone and colleagues (2002) developed a novel FIT recognition task in which examinees are shown 30 items on a sheet, which include the 15 actual items and 15 foils. Due to documentation in adult PVT literature of varied sensitivity and specificity from the various recommended cut scores for the FIT recall task (Strauss et al., 2006), the FIT recognition task was designed to improve classification accuracy. Green and colleagues (2014) found that utilization of a FIT recall cut-score (< 9) yielded very poor sensitivity (.12), but excellent specificity (.98) in a child and adolescent mTBI sample. Whereas, the combination cut-score (< 26) had the best combined score (sensitivity = .55, specificity = .91).

***Wide Range Assessment of Memory and Learning – Second Edition.*** Multiple tasks assessing immediate and delayed memory and learning in visual and verbal formats are included in the WRAML-2. The verbal learning task assesses a child's ability to learn, retain, and recall words from a list. A word list is presented multiple times, and the child or adolescent is asked to freely recall words from the list and then recognize the target words when verbally presented a list that includes target words and foils. To date, no studies have investigated possible WRAML-2 embedded validity measures.

Embedded measures investigated using other verbal learning tasks include a recognition discriminability (RD) index, d prime (*d'*), and forced choice recognition. Investigation of the California Verbal Learning Test –Children's Version (CVLT-C; Delis, Kramer, Kaplan, & Ober, 1994) RD index (assesses ability to distinguish target words from distracter words) indicated variability in recommended cut-off scores that

range from an age-corrected z-score of -0.5 in a mTBI sample (Baker et al., 2014) to age-corrected z-score of -3.0 in a mixed clinical group (Brooks & Ploetz, 2015). In adult PVT literature, d prime, has served as an embedded PVT in the California Verbal Learning Test-Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) recognition subtest (e.g., Curtis, Greve, Bianchini, & Brennan, 2006; Wolfe et al., 2010), but it has not yet been investigated in youth samples.

For the purposes of this study, RD (i.e., ability to discriminate target words from distractors) and response bias (or tendency to acquiesce or nay-say) were calculated based upon descriptions provided in the CVLT-C manual (Delis et al., 2000). A parametric signal detection statistic, d prime (z score that reflects the absolute difference between hit rate and false positive rate) was also calculated (Macmillan & Creelman, 1991). Utilizing the recognition trial from the WRAML-2 Verbal Learning Task the raw number of hits, false alarms, misses, and correct rejections were identified for each participant. Then the hit rate 'H' (i.e., the proportion of YES trials to which a participant responded yes) and false alarm rate "F" (i.e., the proportion of NO trials to which a participant responded YES) were also calculated. Lastly, an excel calculator was used to determine the d prime statistic [$d' = z(H) - z(F)$] which is the difference between the z-transformations of the hit rate 'H' and false alarm rate 'F'.

Recently, an experimental forced-choice recognition task (FCR-C; Lichtenstein, Erdodi, & Linnea, 2017) was developed for the CVLT-C. Recommended cut-off scores varied in sensitivity (sn) and specificity (sp) ($\leq 12$, sn = .14, sp = .97; $\leq 13$, sn = .15, sp = .94; $\leq 14/15$ sn = .31, sp = .87) in this intial CVLT-C FCR-C study with children and adolescents (6 to 15 years of age) with mixed clinical conditions. A novel WRAML-2

Verbal Learning Forced Choice task was created for the purposes of this study (see Appendix G). Item content for foil type was based upon CVLT-II example. Children 8 years and younger learn a 13-item word list, while children 9 years and older learn a 16 item word list, thus forced choice length is 13 or 16 items dependent on age.

***Victoria Symptom Validity Test.*** The Victoria Symptom Validity Test (VSVT; Slick, et al., 1997) is stand-alone computer-administered, forced-choice PVT that includes 48 items. A 5-digit number is presented, must be retained briefly and then the examinee selects the correct answer from two 5 digit numbers. Items are "easy" (the study number and foil do not share common digits) or "difficult" (foils are identical to the study number except for transposition of two digits). Manual published VSVT cut scores include ranges for likely valid scores (≥ 16 total easy or difficult items correct and ≥ 30 total items correct), questionably valid scores (8 to 15 total easy or difficult items correct and 18 to 29 total items correct), and invalid scores (≤ 7 total easy or difficult items correct and ≤ 17 total items correct). Cut scores are also available for examinee response latency on easy or difficult items (e.g.≤ 1.67 seconds average easy item response latency, ≤ 2.68 difficult item response latency). In a mixed clinical sample study with 6- to 19-year-olds, the majority of children could achieve VSVT scores in the manual-recommended valid range for the total score (95%), easy items (97%), and difficult items (84%; Brooks, 2012).

Additional empirically derived VSVT cut-off scores have been proposed in adult samples. For example, in studies with adults who have intractable epilepsy (≤ 20 total difficult items correct; Grote et al., 2000; Loring, Lee, & Meador, 2005), adults who sustained a mTBI (≤ 22 easy items correct, ≤ 17 difficult items correct, and ≤ 40 total

items correct; Silk-Eglit, Lynch, & McCaffrey, 2016 and ≤ 18 easy items correct, ≤ 16 difficult items correct, ≤ 39 total items correct; Jones, 2013) and undergraduate ADHD simulation samples (≤ 18 total difficult items correct; Frazier, Frazier, Busch, Kerwood, & Demaree, 2008) all suggested cut-off scores that are higher than manual recommendations and demonstrate better classification statistics.

***Behavior Assessment System for Children-Second Edition.*** The Behavior Assessment System for Children-Second Edition (BASC-2; Reynolds & Kamphaus, 2004) is a system of questionnaires that assess emotional and behavioral difficulties in children, adolescents, and young adults. Child and adolescent versions were completed by youth, and the parent/guardian completed a parent-report version. Three validity scales are included that are sensitive to various aspects of response distortion (e.g., intentional dissimulation, lack of motivation to respond truthfully, or poor comprehension). The *F* index, also known as the 'fake bad' scale assesses excessive negativity. The *L* index, or 'faking good' scale, assesses for a response pattern in which the respondent tries to portray themselves in an overly positive or defensive manner. Finally, the *V* index, or validity index serves as a basic check for random and/or careless responding, poor reading comprehension, or uncooperative responding. For example, the child is asked to respond to nonsensical statements such as, "I drink 50 glasses of milk per day."

Alpha coefficients were derived for all BASC-2 scales (see Appendix H). Methodological limitations and measurement-based issues should be considered when interpreting reported alpha coefficients. Alpha coefficients reported are likely to be biased based on limited sample sizes (*n* ranged from 14 to 32) and the dimensionality of

scales. Simulation research suggest a sample size approaching 100 would be necessary to obtain unbiased alpha coefficients for these scales (Yurdugul, 2008). The BASC-2 Technical Manual (Reynolds & Kamphaus, 2004) reports generally acceptable alpha coefficients for substantive scales (Parent-report questionnaires: $\alpha \geq .72$; Child and adolescent self-report $\alpha \geq .67$). The Technical Manual does not report validity scale internal consistency. While alpha coefficients for validity scales are technically below an acceptable range (see Appendix H; e.g., Child Self-Report: F Index $\alpha = .48$; L Index $\alpha = .46$; V Index $\alpha = .05$), these values are not unexpected given the previously noted issues and the fact that validity scales are composed of indications of error variance rather than a substantive theoretical construct. Notably, the internal consistency of BASC-2 validity scales are similar to Personality Assessment Inventory (Morey, 2007) and Minnesota Multiphasic Personality Inventory-2-Restructured Form (Ben-Porath, Graham, & Tellegen, 2009) validity scales.

***Behavior Rating Inventory of Executive Function*.** The Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000) is a frequently utilized parent-report measure to assess working memory, planning, organization, emotional control, and other executive function related behaviors. It includes two validity scales: Negativity and Inconsistency. The Negativity scale sums specific items endorsed as "almost always" to determine if the respondent provided an excessively negative or infrequent profile. The Inconsistency scale is used to identify random or careless responding or poor reading accuracy. These scales were validated in the normative group and selected clinical groups; however, no other external SVT studies investigating these validity scales have been conducted.

The BRIEF Technical Manual (Gioia et al., 2000) does not report internal consistency of the Negativity and Inconsistency scales. The internal consistency of BRIEF validity scales was acceptable (See Appendix I; Negativity $\alpha = .82$; Inconsistency $\alpha = .89$). Internal reliability for index ($\alpha \geq .93$) and sub-scales ($\alpha \geq .78$) was also acceptable, though slightly below reported alpha levels in the Technical Manual ($\alpha \geq .81$).

## Data Analysis

This study was comprised of four groups (control, simulation, community-clinical, and medical center-clinical), thus group similarities in age and estimated intellectual functioning were evaluated through multivariate analysis of variance. Additionally, it was important to determine if the stratification procedure for control and simulation group assignment was successful. Thus multivariate results and demographic characteristics were evaluated for similarity. Also in regard to sample characteristics, the ability of simulation participants to suppress performance on neuropsychological tasks and alter their responses on questionnaires in order to present with impairments was assessed through multivariate analyses. Simulation participants also completed manipulation check questionnaires, thus descriptive statistics were generated to evaluate instruction comprehension and adherence with children and parents.

For the control and clinical groups, correlations of age, estimated intellectual functioning, and memory with PVTs are presented. Proposed PVT cut-off scores are available for VSVT, DS ACSS, RDS, RDS-R, Rey FIT, and AST from either adult or pediatric samples. The percentage of children and adolescents from the control and

clinical groups that can "pass" these PVT cut-offs are presented. Additionally, PVT "passing" rates for younger children and those with more significant cognitive impairment are displayed (i.e., FSIQ < 70).

Receiver Operating Curve (ROC) analyses were conducted and classification statistics (e.g., sensitivity, specificity, predictive power) were generated for PVTs and SVTs that differentiated simulation participants from control and clinical participants in multivariate analysis and display acceptable AUC values. Subsequently, optimal PVT and SVT cut-off scores were identified. Lastly, correlations between PVTs and SVTs were generated to evaluate the degree to which measures converge.

## Results

### Multivariate Analysis of Demographics

Consistent with Aim 1, multivariate analyses were conducted to evaluate differences between the control, simulation, community-clinical, and medical center-clinical groups in demographic characteristics and performance and symptom validity test results. Given multiple comparisons, Bonferroni correction was applied in each model. Levene's test of homogeneity of variance was frequently violated, thus a more conservative alpha level ($p < .01$) was frequently utilized (Tabachnick & Fidell, 2013).

In regard to demographic characteristics (see Table 2), multivariate analysis indicated that groups differed in age and years of education [Wilk's $\Lambda = .779$, $F(6, 370) = 8.21$, $p < .001$, partial $\eta^2 = .12$]. The medical center-clinical group was significantly older [$F(3, 186) = 16.62$, $p < .001$, partial $\eta^2 = .21$] and in a higher grade [$F(3, 186) = 14.17$, $p$

< .001, partial $\eta^2 = .19$] than the community-recruited control and clinical groups but was not significantly older than the simulation group. The community-recruited control and simulation groups did not significantly differ in age or years of education; thus, stratification on age was successful. Additionally, the community-recruited simulation and control groups and the medical-center clinical group included an approximately equal number of males and females.

Table 2
*Participant Demographics and MANOVA Results*

| | Community | | | MCW | | |
|---|---|---|---|---|---|---|
| | Control[a] *n* = 23 | Simulation[b] *n* = 20 | Clinical[c] *n* = 23 | Clinical[d] *n* = 125 | *F* | Post hoc |
| Age (years.months) | 9.82 (2.10) | 11.72 (2.62) | 10.25 (2.52) | 13.06 (2.55) | 16.62*** | (a=b=c); (b = d) (a=c) < d |
| Age range | 8.02 - 15.04 | 8.01 - 16.30 | 8.00 - 16.11 | 7.10 - 16.11 | | |
| Grade in school | 4.39 (2.29) | 6.35 (2.60) | 4.78 (2.78) | 7.47 (2.58) | 14.17*** | (a=b=c); (a=c) < d |
| % female | 47.8% (11) | 50% (10) | 39.1% (9) | 51.2% (64) | | |
| % Caucasian | 56.5% (13) | 85% (17) | 69.6% (16) | 52% (65) | | |
| WISC-V Vocabulary (ss) | 11.57 (3.04) | 12.60 (3.07) | 10.74 (4.01) | 8.58 (2.94) | 14.40*** | d < (a=b=c) |
| WISC-V Matrix Reasoning (ss) | 10.78 (3.18) | 11.00 (3.34) | 9.00 (3.29) | 8.73 (3.10) | 4.98** | d < (a = b); (c= d); (a=b=c) |
| WISC-V Digit Span (ss) | 10.91 (2.07) | 5.85 (4.72) | 9.17 (3.23) | 8.30 (2.85) | 9.90*** | b < (c = d) < a |
| Bracken (Raw) | 18 (0) | 18 (0) | 18 (0) | - | | |
| WRAML-2 | | | | | | |
| Immediate Recall (ss) | 10.18 (3.26) | 6.60 (3.12) | 10.52 (3.41) | 9.07 (2.52) | 7.56*** | b < (a=c=d) |
| Delayed Recall (ss) | 9.77 (3.88) | 7.00 (3.69) | 10.74 (2.83) | 9.30 (2.25) | 5.95** | b < (a =c) |

*Note:* * = $p < .05$; ** = $p < .01$; *** = $p < .001$; WISC-V = Wechsler Intelligence Scale for Children – Fifth Edition; ss = scaled score; WRAML-2 = Wide Range Assessment of Memory and Learning – 2nd Edition.

WISC-V Vocabulary and Matrix Reasoning subtests were utilized to estimate cognitive ability. Of note, participants assigned to the simulation condition completed Vocabulary, Matrix Reasoning, and Bracken number identification subtests prior to receiving simulation instructions. Community recruited and medical center groups differed in performance on the WISC-V Vocabulary and Matrix Reasoning subtests [Wilk's $\Lambda = .791$, $F(6, 354) = 7.36$, $p < .001$, partial $\eta^2 = .11$]. The medical center-clinical participants had significantly lower WISC-V Vocabulary subtests scores than the community recruited groups [$F(3, 178) = 14.40$, $p < .001$, partial $\eta^2 = .20$], though, notably, group means were within the average range. The medical center-clinical

participants also had significantly lower Matrix Reasoning subtest scores [$F(3, 178) = 4.98$, $p < .01$, partial $\eta^2 = .08$] than the community recruited control and simulation groups; however they had similar performance to the community recruited-clinical participants. Mean performance on Matrix Reasoning was also in the average range for all groups. Importantly, on estimates of intellectual functioning (see Table 2), simulation participants did not differ from control participants in word knowledge [$t(41) = -1.11$, $p = .27$] or nonverbal reasoning [$t(41) = -0.22$, $p = .83$], given that the Vocabulary and Matrix Reasoning subtests were administered prior to simulation instructions.

**Analysis of Simulation Adherence**

Also within Aim 1, assessment of simulation adherence was completed. Performance of children and adolescents assigned to simulate was evaluated to determine if they could intentionally suppress performance or alter responses. Following simulation instructions, participants performed more poorly than clinical and control participants on the WISC-V Digit Span [$F(3, 183) = 9.90$, $p < .001$] task. Simulators also suppressed memory performance (see Table 2) to be lower than control and clinical participants [Wilk's $\Lambda = .82$, $F(6, 228) = 3.97$, $p < .001$]. The simulation group also had poorer immediate recall of words from the WRAML-2 Verbal Learning task than the control and clinical groups [$F(3, 115) = 7.56$, $p < .001$, partial $\eta^2 = .17$]. Simulation participants also had poorer delayed recall of words than the control and community recruited-clinical groups but not the medical center-clinical group [$F(3, 115) = 5.95$, $p < .01$, partial $\eta^2 = .13$] after correcting for multiple comparisons.

In regard to symptom validity scales, child and adolescent simulators altered their response pattern on the BASC-2 which elevated the V Index [$F(2, 62) = 5.71$, $p < .01$, partial $\eta^2 = .16$] but not the other validity scales. Evaluation of non-symptom validity scales on the BASC-2 indicated that simulators did not differ from control or clinical participants on BASC-2 clinical scales [Wilk's $\Lambda = .38$, $F(38, 88) = 1.44$, $p = .09$]. Therefore, the request that simulators exaggerate cognitive symptoms associated with a brain injury did not cause them to exaggerate self-reported psychological difficulties.

Parents/guardians assigned to the simulation condition also altered their response pattern on parent questionnaires which elevated the BASC-2 F Index [$F(2, 63) = 8.64$, $p < .01$, partial $\eta^2 = .22$] and BRIEF Negativity scale [$F(2, 63) = 14.56$, $p < .01$, partial $\eta^2 = .32$]. Of note, parents/guardians also altered responses in a manner that resulted in scale elevation on all BASC-2 clinical [Wilk's $\Lambda = .25$, $F(36, 92) = 2.58$, $p < .001$, partial $\eta^2 = .50$] and BRIEF clinical [Wilk's $\Lambda = .35$, $F(22, 106) = 3.31$, $p < .001$, partial $\eta^2 = .41$] scales. Clinical scales from the simulation group were significantly different from control and clinical participants in nearly every scale (see Table 3).

Table 3
*Parent-report BASC-2 and BRIEF Clinical Scale Manova Results (T Scores)*

| | Control[a] *n* = 23 | Simulation[b] *n* = 20 | Clinical[c] *n* = 23 | *F* | Post hoc |
|---|---|---|---|---|---|
| BASC-2 | | | | | |
| Composite Scores | | | | | |
| Externalizing Problems | 48.48 (9.61) | *67.90 (19.68)* | 54.35 (12.41) | 10.32*** | (a = c) < b |
| Internalizing Problems | 48.35 (10.85) | **76.75 (17.62)** | 56.96 (13.42) | 22.64*** | (a = c) < b |
| Behavioral Symptoms | 47.43 (8.36) | **77.65 (22.24)** | 59.09 (10.55) | 23.19*** | a < c < b |
| Adaptive Skills | 52.52 (8.96) | 28.35 (16.08) | 42.00 (9.70) | 22.53*** | a < c < b |
| Scale Scores | | | | | |
| Hyperactivity | 49.96 (9.42) | *69.55 (18.01)* | 57.43 (14.21) | 13.70*** | a < c < b |
| Aggression | 50.61 (10.35) | *62.20 (17.59)* | 51.57 (10.55) | 5.11** | (a = c) < b |
| Conduct Problems | 48.35 (9.23) | *66.45 (19.09)* | 52.83 (11.27) | 10.16*** | (a = c) < b |
| Anxiety | 48.48 (12.34) | *64.90 (8.60)* | 55.91 (13.67) | 10.26*** | (a = c) < b |
| Depression | 49.74 (9.08) | **74.55 (21.00)** | 55.96 (9.17) | 18.34*** | (a = c) < b |
| Somatization | 47.74 (11.25) | **75.70 (18.89)** | 55.13 (15.51) | 18.73*** | (a = c) < b |
| Atypicality | 45.57 (3.87) | **80.05 (23.84)** | 59.65 (12.57) | 27.57*** | a < c < b |
| Withdrawal | 47.52 (9.64) | **72.15 (18.40)** | 58.26 (13.33) | 16.52*** | a < c < b |
| Attention | 47.78 (9.58) | *69.40 (13.19)* | 59.70 (8.67) | 22.80*** | a < c < b |
| Adaptability | 51.65 (9.38) | *31.65 (14.27)* | 40.78 (10.17) | 16.87*** | b < c < a |
| Social Skills | 49.91 (10.36) | *35.25 (15.13)* | 46.26 (11.40) | 8.07*** | b < (a = c) |
| Leadership | 54.65 (9.96) | *35.15 (13.17* | 45.78 (10.53) | 16.19*** | b < c < a |
| Daily Living | 50.91 (9.36) | **29.25 (13.27)** | 41.57 (9.71) | 21.56*** | b < c < a |
| Communication | 53.70 (7.31) | **25.65 (15.85)** | 41.70 (10.80) | 31.19*** | b < c < a |
| BRIEF | | | | | |
| Inhibit | 46.78 (8.71) | *65.95 (16.31)* | 56.78 (11.58) | 12.85*** | a < c < b |
| Shift | 47.65 (7.11) | **72.85 (18.96)** | 63.13 (14.37) | 17.67*** | a < c < b |
| Emotional Control | 48.39 (9.02) | *64.90 (16.85)* | 57.13 (9.56) | 10.05*** | (a = c) < b |
| BRI | 47.39 (7.97) | *69.70 (18.39)* | 60.09 (10.53) | 16.59*** | a < c < b |
| Initiate | 48.13 (7.52) | *68.70 (15.65)* | 59.13 (11.06) | 16.72*** | a < c < b |
| Working Memory | 50.09 (9.16) | **74.15 (15.16)** | 63.22 (10.54) | 22.74*** | a < c < b |
| Planning | 46.35 (8.15) | **70.95 (14.11)** | 57.87 (9.12) | 28.84*** | a < c < b |
| Organization of Materials | 51.96 (8.76) | 60.30 (11.06) | 53.52 (8.67) | 4.59* | (a = c) < b |
| Monitor | 45.26 (8.32) | *65.55 (14.08)* | 57.35 (11.59) | 17.22*** | a < (b = c) |
| MI | 48.04 (7.79) | **70.90 (15.41)** | 59.78 (9.24) | 22.87*** | a < c < b |
| GEC | 47.78 (6.74) | **72.10 (16.59)** | 60.74 (10.04) | 23.74*** | a < c < b |

*Note:* * = $p < .05$; ** = $p < .01$; *** = $p < .001$; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function; Scores are displayed in T-scores. Bolded scores are considered clinically significant; Italicized scores are at risk, but sub-clinical.

Finally, a manipulation check was conducted using a 5-pointlikert scale (see Appendix E) that was administered at the end of the testing session. Children and adolescents in the simulation group reported strong comprehension ($M$ = 4.05, SD = 1.03) of simulation instructions and moderate ability to follow the instructions ($M$ = 3.47, SD = 1.02). Similarly, parents assigned to simulate indicated nearly complete understanding of instructions ($M$ = 4.60, SD = 0.68) and moderate to strong ability to follow instructions ($M$ = 3.90, SD = 1.52) while completing questionnaires.

**Correlations of Age, IQ, and Memory with PVTs & SVTs**

In accordance with Aim 1, relationships between performance and symptom validity test results and age, estimates of intellectual functioning, and memory were evaluated for children and adolescents in the control and clinical groups ($n$ = 171; see Table 4). Correlations for the control and clinical were first investigated separately and then, due to similarity, were combined in Table 4. Additionally, given the small control group sample size ($n$ = 23) it was determined that correlations would be more stable if the medical center- and community-clinical groups were combined with the control group. VSVT scores were generally associated with age and estimates of intellectual functioning but generally not related to memory. Performance on the VSVT was weakly to moderately, positively associated with age and WISC-V Vocabulary. VSVT performance was also moderately, positively related to WISC-V Matrix Reasoning performance. VSVT total easy, difficult, and correct items were moderately, positively associated with WISC-V Digit Span. The easy latency time displayed a small, negative association with Digit Span. Additionally, immediate recall performance on the WRAML-2 word list was weakly, positively associated with the number of VSVT Difficult Items correct score.

Table 4
*Correlations of PVTs and SVTs with Age, Intellectual Estimates, and Memory for Control and Clinical Participants*

| | Age | WISC-V Vocab | WISC-V Matrix Reasoning | WISC-V Digit Span | WRAML-2 Immediate Recall | WRAML-2 Delayed Recall |
|---|---|---|---|---|---|---|
| VSVT (*n* = 171) | | | | | | |
| Easy Correct | .25*** | .31*** | .40*** | .40*** | .15 | .12 |
| Difficult Correct | .30*** | .29*** | .45*** | .38*** | .20* | .18 |
| Total Correct | .30*** | .32*** | .47*** | .42*** | .19 | .17 |
| Easy Latency[1] | -.46*** | -.08 | -.18* | -.16* | -.12 | .03 |
| Difficult Latency[1] | -.29*** | .01 | -.04 | -.10 | .02 | .15 |
| Digit Span (*n* = 167) | | | | | | |
| ACSS | -.03 | .53*** | .50*** | -- | .40*** | .26* |
| RDS | .23** | .40*** | .42*** | -- | .36*** | .21* |
| RDS-R | .29*** | .44*** | .43*** | -- | .39*** | .23* |
| WRAML-2 (*n* = 100) | | | | | | |
| Signal Detection | | | | | | |
| Discriminability | .12 | .26* | .11 | .28** | .49*** | .33** |
| Response Bias | -.23* | .01 | -.02 | .01 | -.05 | -.10 |
| d Prime | .22* | .24* | .15 | .29** | .50*** | .35*** |
| Forced Choice | | | | | | |
| Raw Correct | .75*** | .18 | -.01 | .07 | .11 | .13 |
| Percent Correct | .01 | -.12 | -.05 | -.05 | .11 | .10 |
| Rey FIT (*n* = 45) | | | | | | |
| Recall Correct | .33* | .39** | .43** | .17 | .28 | .26 |
| Recognition | .31* | .33* | .30* | .12 | .30* | .25 |
| False Positives | .08 | -.18 | -.21 | .01 | -.16 | .01 |
| Combination Score | .25 | .35* | .34* | .11 | .24 | .22 |
| AST (*n* = 45) | | | | | | |
| Alphabet[1] | -.23 | -.10 | -.30* | -.11 | .03 | -.13 |
| Alphabet Errors | -.08 | -.01 | -.27 | -.12 | -.27 | -.05 |
| Counting[1] | -.30* | -.33* | -.40** | -.46** | -.02 | -.04 |
| Counting Errors | -.21 | -.12 | -.17 | -.12 | -.09 | -.05 |
| Days of Week[1] | -.31* | -.35* | -.32* | -.41** | -.10 | -.27 |
| Days of Week Errors | -.15 | -.36* | -.22 | -.36* | -.23 | -.27 |
| Months[1] | -.35* | -.25 | -.30* | -.09 | -.06 | -.14 |
| Months Errors | -.16 | -.31* | -.28 | -.24 | .02 | -.02 |
| Total[1] | -.36* | -.25 | -.37* | -.19 | -.04 | -.16 |
| Total Errors | -.23 | -.25 | -.41** | -.29 | -.26 | -.11 |
| BASC-2 (*n* = 45) | | | | | | |
| F Index | -.12 | -.32* | -.34* | -.19 | -.14 | -.24 |
| Consistency | -.15 | -.32* | -.23 | -.01 | -.03 | .01 |
| L Index | -.22 | -.39* | -.44** | -.40** | -.44** | -.27 |
| V Index | -.26 | -.10 | -.17 | -.14 | -.09 | -.07 |

*Note:* * = $p < .05$; ** = $p < .01$; *** = $p < .001$; VSVT = Victoria Symptom Validity Test; Dif. = Difficult; [1] = time in seconds; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task; BASC-2 = Behavior Assessment System for Children – Second Edition.

Embedded validity tests from the WISC-V Digit Span test were weakly, positively associated with age and delayed verbal memory, while moderately, positively associated with WISC-V Vocabulary and Matrix Reasoning and immediate verbal learning recall (see Table 4).

Embedded validity scales on the WRAML-2 were somewhat associated with age and word knowledge and WRAML-2 recall conditions, though there was variability across measures. The Forced Choice task raw score was strongly, positively related to age, though notably, children and adolescents are administered a different number of items based on their age (i.e., 8 years and younger 13 items; 9 years and older 16 items). Percentage correct on the Forced Choice task was not associated with age. Recognition discriminability was weakly, positively associated with WISC-V Vocabulary and Digit Span, while moderately, positively related to WRAML-2 immediate and delayed recall conditions. Response bias was weakly, negatively associated with age. D prime was weakly, positively correlated with age and WISC-V Vocabulary, while moderately, positively correlated with WISC-V Digit Span and WRAML-2 immediate and delayed recall conditions.

The Rey FIT tasks were generally associated with age and intellectual estimates. Rey FIT immediate recall was moderately, positively associated with age and WISC-V Vocabulary and Matrix Reasoning tasks. The Rey FIT recognition condition was moderately, positively associated with age, WISC-V Vocabulary and Matrix Reasoning, and WRAML-2 Immediate Recall. The Rey FIT combination score, which accounts for

recall and errors, was moderately, positively associated with WISC-V Vocabulary and Matrix Reasoning.

In regard to the Automatized Sequences Task, completion time was generally associated with age and estimates of intellectual functioning, though there were variations across conditions. Time to say the alphabet was moderately, negatively associated with WISC-V Matrix Reasoning. Time to count to 20 and state the days of the week were moderately, negatively associated with age, and WISC-V Vocabulary, Matrix Reasoning, and Digit Span. Time to say the months of the year and total completion time were moderately, negatively associated with age and WISC-V Matrix Reasoning. The amount of errors on counting and months of the year were moderately, negatively associated with WISC-V Vocabulary; while the total number of errors across conditions was moderately, negatively correlated with WISC-V Matrix Reasoning.

BASC-2 validity scales were generally negatively associated with word knowledge performance, though there was variability. The F Index was moderately, negatively associated with WISC-V Vocabulary and Matrix Reasoning. Consistency in responding was moderately, negatively related to WISC-V Vocabulary. The L Index was negatively, moderately correlated with WISC-V Vocabulary, Matrix Reasoning, and Digit Span, and WRAML-2 Immediate Recall. The V Index was not associated with age, intellectual estimates, working memory, or verbal memory tasks.

**Frequency of Invalidity**

For children and adolescents in the control and clinical groups, PVT and SVT results were investigated to determine what percentage of participants performed below

previously established cut-off scores to identify invalidity, which were derived from adult or pediatric mTBI samples. Additionally, the frequency of performance below established cut-off scores was evaluated in two clinical sub-samples a) children 10 years or younger (*n* = 58) and b) children with possible Intellectual Disability (ID; *n* = 14; see Table 5). Individuals assigned to the ID group either received a formal diagnosis of ID, had a Full Scale IQ of less 70, or obtained scaled scores of < 5 on both WISC-V Vocabulary and Matrix Reasoning subtests. Due to relatively restricted sample sizes, the two clinical sub-samples include individuals from both the community and MCW clinical groups.

Table 5
*Percentage of Control and Clinical Participants at Previously Identified Cut-off Scores*

| | Cut-off | Control | Community Clinical | MCW Clinical | ≤ 10 years old | Intellectual Disability (ID)[g] |
|---|---|---|---|---|---|---|
| | | *n* = 23 | *n* = 23 | *n* = 125 | *n* = 58 | *n* = 14 |
| VSVT[a] | | | | | | |
| Invalid | | | | | | |
| Easy Correct | ≤ 7 | 0% (0) | 0% (0) | 0% (0) | 0% (0) | 0% (0) |
| Difficult Correct | ≤ 7 | 0% (0) | 0% (0) | 0% (0) | 0% (0) | 0% (0) |
| Total Correct | ≤ 17 | 0% (0) | 0% (0) | 0% (0) | 0% (0) | 0% (0) |
| Questionable | | | | | | |
| Easy Correct | 8 - 15 | 0% (0) | 4% (1) | 5% (6) | 9% (5) | 29% (4) |
| Difficult Correct | 8 - 15 | 5% (1) | 22% (5) | 18% (23) | 26% (15) | 57% (8) |
| Total Correct | 18 - 29 | 0% (0) | 9% (2) | 6% (7) | 11% (6) | 36% (5) |
| Valid | | | | | | |
| Easy Correct | ≥ 16 | 100% (22) | 96% (22) | 95% (119) | 91% (52) | 71% (10) |
| Difficult Correct | ≥ 16 | 95% (21) | 78% (18) | 81% (102) | 76% (42) | 43% (6) |
| Total Correct | ≥ 30 | 100% (22) | 91% (21) | 94% (118) | 89% (51) | 64% (9) |
| Digit Span | | | | | | |
| ACSS[b] | ≤ 5 | 0% (0) | 26% (6) | 13% (16) | 15% (8) | 64% (9) |
| RDS[b] | ≤ 6 | 5% (1) | 17% (4) | 20% (24) | 24% (13) | 50% (7) |
| RDS-R[c] | ≤ 11 | 14% (3) | 39% (9) | 39% (39) | 32% (10) | 57% (8) |
| Rey FIT[d] | | | | | | |
| Recall Correct[d] | ≤ 7 | 5% (1) | 13% (3) | -- | 13% (4) | 0% (0) |
| Combination[e] | < 26 | 41% (9) | 35% (8) | -- | 52% (16) | 50% (1) |
| AST[f] | | | | | | |
| Alphabet[1] | ≥ 8 | 18% (4) | 26% (6) | -- | 28% (9) | 0% (0) |
| Counting[1] | ≥ 6 | 23% (5) | 35% (8) | -- | 38% (12) | 0% (0) |
| Days of Week[1] | ≥ 4 | 14% (3) | 22% (5) | -- | 22% (7) | 0% (0) |
| Months[1] | ≥ 10 | 18% (4) | 48% (11) | -- | 41% (13) | 0% (0) |
| Total[1] | ≥ 27 | 18% (4) | 39% (9) | -- | 38% (12) | 0% (0) |

*Note:* Samples size varied for some measures since some tasks were only administered to community participants or due to missing data so both percentage and frequencies are reported. Cut-off scores were identified from the following publications; VSVT[a] = Slick et al., 1997 VSVT manual with adult normative data; ACSS[b] and RDS[b] = Kirkwood et al., 2011 child and adolescent mTBI sample; RDS-R[c] = Spencer et al. 2013 and Young et al., 2012 adult veterans; Rey FIT Recall Correct[d] = Blaskewitz et al., 2008 child simulation study; Rey FIT Recognition[e] = Green et al., 2014 child and adolescent mTBI sample; AST[f] = Kirkwood et al., 2014 child and adolescent mTBI sample; [1] = time in seconds

No children or adolescents in the control and clinical groups exhibited invalid performance on the VSVT according to manual published cut-off scores. However, children and adolescents did exhibit VSVT performance in the "questionable" range. Additionally, children and adolescents were more likely to be in the "questionable" range if they were in one of the clinical groups, were younger or had an intellectual disability.

Using cut-off scores derived from pediatric mTBI samples, children and adolescents in the community and medical-center clinical groups were more likely to be classified as providing invalid performance on the Digit ACSS and RDS than controls. Additionally, younger children and individuals with intellectual disability were more likely to have invalid Digit ACSS and RDS scores. Cut-off scores for the RDS-R have not yet been established for a child and adolescent sample. When using the adult veteran derived RDS-R cut-off score, a third of children and adolescents would be within an invalid group and, notably, half of children with an intellectual disability would be within an invalid group.

The Rey FIT and Automatized Sequences Task were only completed by the community recruited control and clinical participants ($n$ = 46). A child and adolescent cut-off score for the Rey FIT Recall Correct Score had been previously established in a simulation study with only control and simulation participants. The majority of children in the present study could pass the previously established cut-off score, though the only children within the invalid range were 10 years or younger. A child and adolescent cut-off score for the Rey FIT Combination Score was established using patients with mTBI. When applying that cut-off score to the current mixed clinical and control samples, over a third of children and adolescents would be classified as providing invalid performance, while half of children 10 and younger and individuals with intellectual disability would be within the invalid range.

Cut-off scores for the AST were established from a child and adolescent mTBI sample. Approximately 15-20% of control participants would be identified as providing invalid performance if the previously established cut-off scores were utilized. Further, 20

to 50% of clinical participants would be identified as providing invalid performance. Additionally, younger children were more likely to exhibit performance in the invalid range. However, the two children with intellectual disabilities that completed the AST were able to perform within the previously established valid range.

In summary, these results supported the hypothesis that a majority of children in the control and clinical groups would "pass" probabilistic measures of performance validity evidenced by the VSVT results. Additionally, the hypothesis that younger children (e.g., ≤ 10 years) and children with greater degrees of cognitive impairment (e.g., intellectual disability) would demonstrate greater rates of PVT failure on embedded PVTs was supported.

**Performance Validity Multivariate Analyses**

PVT performance differences were evaluated for simulator, control, and clinical groups in accordance with Aim 1 (see Table 6). Multivariate analysis indicated that groups differed in performance on the VSVT measures [Wilk's $\Lambda = .84$, $F(15, 500) = 2.17$, $p < .01$, partial $\eta^2 = .06$]. The simulation group had significantly fewer VSVT easy items correct than the control or clinical groups [$F(3, 185) = 6.77$, $p < .001$, partial $\eta^2 = .10$]. The simulation group also had significantly fewer VSVT difficult items correct than the control and medical center-clinical groups but not the community-clinical group [$F(3, 185) = 4.84$, $p < .01$, partial $\eta^2 = .07$]. Similarly, the simulation group had fewer VSVT total items correct than the control and medical center-clinical groups but not the community-clinical group [$F(3, 185) = 6.25$, $p < .001$, partial $\eta^2 = .09$]. While VSVT easy item latency appears significantly different across groups, the more conservative

alpha level requirement was not met [$F(3, 185) = 2.66$, $p < .05$, partial $\eta^2 = .04$]. Groups did not differ in VSVT difficult item latency [$F(3, 185) = 1.08$, $p = .36$, partial $\eta^2 = .01$].

Table 6
*Performance Validity Test Descriptive Statistics and MANOVA Results*

| | Community | | | MCW | | |
|---|---|---|---|---|---|---|
| | Control[a] *n* = 23 | Simulation[b] *n* = 20 | Clinical[c] *n* = 23 | Clinical[d] *n* = 125 | *F* | Post hoc |
| *VSVT* | | | | | | |
| Easy Correct | 23.50 (1.06) | 19.53 (5.23) | 22.83 (2.21) | 22.45 (2.95) | 6.77*** | b < (a=c=d) |
| Dif. Correct | 21.64 (2.54) | 16.95 (5.40) | 19.04 (4.85) | 19.95 (3.95) | 4.84** | b < (a = d) b = c |
| Total Correct | 45.14 (3.41) | 36.47 (9.92) | 41.87 (6.62) | 42.38 (6.38) | 6.25*** | b < (a = d) b = c |
| Easy Latency[1] | 2.65 (1.24) | 3.12 (2.61) | 2.93 (1.30) | 2.36 (1.05) | 2.66* | |
| Dif. Latency[1] | 4.13 (1.45) | 4.28 (2.53) | 4.26 (1.91) | 3.76 (1.52) | 1.08 | |
| *Digit Span* | | | | | | |
| DS ACSS | 10.91 (2.07) | 5.85 (4.72) | 9.17 (3.23) | 8.30 (2.75) | 10.16*** | b < d < (a=c) |
| RDS | 8.23 (1.07) | 5.90 (2.51) | 7.96 (1.94) | 7.87 (1.73) | 7.74*** | b < (a=c=d) |
| RDS-R | 12.91 (1.48) | 9.30 (3.81) | 11.91 (3.15) | 12.03 (2.79) | 6.46*** | b < (a=c=d) |
| *WRAML-2 Verbal Learning* | | | | | | |
| Signal Detection Measures | | | | | | |
| Response Bias | .05 (.24) | -.04 (.32) | .12 (.30) | .07 (.30) | 1.17 | |
| Discriminability | 95.31 (3.46) | 75.52 (22.33) | 94.32 (5.72) | 91.99 (11.42) | 11.80*** | b < (a=c=d) |
| d Prime | 3.66 (.61) | 2.04 (2.06) | 3.63 (.84) | 3.46 (1.17) | 8.24*** | b < (a=c=d) |
| Forced Choice | | | | | | |
| Raw Score | 14.73 (1.49) | 11.60 (4.62) | 14.70 (1.52 | -- | 8.28** | b < (a=c) |
| Percent Correct | 99.68 (1.49) | 74.36 (30) | 100 (0.0) | -- | 16.25*** | b < (a=c) |
| *Rey FIT* | | | | | | |
| Recall Correct | 13.36 (2.50) | 11.25 (4.67) | 12.04 (3.88) | -- | 1.71 | |
| Recognition | 12.82 (2.26) | 10.00 (4.86) | 12.30 (4.08) | -- | 3.14* | b < (a = c) |
| False Positives | 0.32 (0.89) | 1.70 (2.81) | 0.35 (0.71) | -- | 4.50* | (a = c) < b |
| Combination Score | 25.86 (4.45) | 20.05 (9.97) | 23.39 (8.61) | -- | 2.80 | |
| *AST* | | | | | | |
| Alphabet[1] | 6.32 (3.46) | 15.80 (10.83) | 8.17 (8.33) | -- | 8.10*** | (a = c) < b |
| Alphabet[2] | 0.05 (0.21) | 1.20 (1.82) | 1.09 (3.37) | -- | 1.73 | |
| Counting[1] | 4.68 (1.13) | 12.65 (11.58) | 5.70 (2.67) | -- | 8.85*** | (a = c) < b |
| Counting[2] | 0.05 (0.21) | 0.65 (1.50) | 0.26 (0.92) | -- | 1.96 | |
| Days of Week[1] | 2.45 (0.80) | 5.70 (3.96) | 3.00 (1.48) | -- | 10.79*** | (a = c) < b |
| Days of Week[2] | 0.14 (0.35) | 0.55 (1.15) | 0.39 (0.58) | -- | 1.63 | |
| Months[1] | 8.00 (8.47) | 11.80 (7.10) | 11.48 (9.40) | -- | 1.35 | |
| Months[2] | 0.73 (2.33) | 1.20 (1.74) | 1.17 (1.47) | -- | 0.44 | |
| Total[1] | 21.45 (12.02) | 46.30 (30.37) | 28.35 (18.21) | -- | 7.60*** | (a = c) < b |
| Total (Err.) | 1.09 (2.43) | 3.60 (5.39) | 2.91 (3.55) | -- | 2.33 | |

*Note:* **p* < .05; ***p* < .01; ****p* < .001; VSVT = Victoria Symptom Validity Test; Dif. = Difficult; [1] = time in seconds; [2] = Errors; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task

Groups also differed in performance on embedded Digit Span validity measures [Wilk's $\Lambda$ = .706, $F(9, 387) = 6.62$, $p < .001$, partial $\eta^2 = .11$]. The simulation group had a significantly lower Digit Span Age Corrected Scaled Score (DS ACSS) than all other groups, though the medical center-clinical group had a significantly lower score than community-clinical and control groups [$F(3, 161) = 10.16$, $p < .001$, partial $\eta^2 = .16$]. The simulation group had a significantly lower RDS score than clinical and control groups [$F(3, 161) = 7.74$, $p < .001$, partial $\eta^2 = .13$]. Similarly, the simulation group had a significantly lower RDS-R score than clinical and control groups [$F(3, 161) = 6.46$, $p < .001$, partial $\eta^2 = .12$].

Some of the medical center-clinical group ($n = 54$) completed the WRAML-2 Verbal Learning task; whereas all of the community-recruited control, simulation, and clinical groups completed the verbal memory task and an experimental Forced Choice recall task. Response bias, recognition discriminability, and d prime indices were generated (Macmillan & Creelman, 1991). Two multivariate analyses were conducted for the WRAML-2 since the medical center-clinical group did not complete the experimental Forced Choice measure. Groups differed in measures of response bias, discriminability, and d prime [Wilk's $\Lambda$ = .64, $F(9, 250.83) = 5.64$, $p < .001$, partial $\eta^2 = .14$]. Groups did not significantly differ on the response bias measure [$F(3, 105) = 1.17$, $p = .33$, partial $\eta^2 = .03$]. The simulation group performed more poorly than control or clinical groups on recognition discriminability [$F(3, 105) = 11.80$, $p < .001$, partial $\eta^2 = .25$] and d prime [$F(3, 105) = 8.24$, $p < .001$, partial $\eta^2 = .19$].

In the second multivariate analysis that included only the community-recruited control, simulation, and clinical groups, groups differed on WRAML-2 Verbal Learning

Forced Choice performance [Wilk's $\Lambda$ = .63, $F(4, 120) = 8.82$, $p < .001$, partial $\eta^2 = .21$]. The simulation group had fewer items correct on the Forced Choice task [Forced Choice total correct $F(2, 62) = 8.28$, $p < .01$, partial $\eta^2 = .21$; Forced Choice percent correct $F(2, 62) = 16.25$, $p < .001$, partial $\eta^2 = .34$].

The medical center-clinical group did not complete the remaining two PVTs; the Rey FIT and AST so multivariate analyses were conducted with only the community-recruited groups. Community recruited groups differed on Rey FIT measures [Wilk's $\Lambda$ = .74, $F(8, 118) = 2.45$, $p < .05$, partial $\eta^2 = .14$]. Groups did not differ in the total number of Rey FIT test stimuli remembered during an immediate recall condition [$F(2, 62) = 1.71$, $p = .19$]. On a delayed recognition task, simulators performed more poorly than clinical or control participants [$F(2, 62) = 3.14$, $p = .05$, partial $\eta^2 = .09$]; however, the alpha was larger than recommended value ($p < .01$) due to Levene's test violation. Also on the delayed recognition task, simulators had more false positive errors than control or clinical participants [$F(2, 62) = 4.50$, $p = .02$, partial $\eta^2 = .13$]; however, the alpha was also larger than the more conservative recommended value given homogeneity of variance violation. Groups did not differ on the Rey FIT test combination score, which accounts for the number of correctly recalled items and recognition errors [$F(2, 62) = 2.80$, $p = .07$].

Completion time for the four AST conditions (e.g., saying the alphabet, counting to 20, saying the days of the week, and saying the months of the year), the total completion time, and error rate across the conditions and total error rate were evaluated and multivariate analysis suggested that groups differed [Wilk's $\Lambda$ = .50, $F(20, 106) = 2.17$, $p < .01$, partial $\eta^2 = .29$]. The simulators took longer to say the alphabet [$F(2, 62) =$

8.10, $p < .001$, partial $\eta^2 = .21$], count to 20 [$F(2, 62) = 8.85$, $p < .001$, partial $\eta^2 = .22$], and say the days of the week [$F(2, 62) = 10.79$, $p < .001$, partial $\eta^2 = .26$] than control or clinical participants. However, groups did not differ in the amount of time it took to say the months of the year [$F(2, 62) = 1.35$, $p = .27$]. When the amount of time was added together for all conditions, simulation participants were slower than control and clinical participants [$F(2, 62) = 7.60$, $p < .001$, partial $\eta^2 = .20$]. Groups did not differ in the number of errors made in any condition [alphabet errors $F(2, 62) = 1.73$, $p = .19$; counting errors $F(2, 62) = 1.96$, $p = .15$; days of the week errors $F(2, 62) = 1.63$, $p = .20$; months of year errors $F(2, 62) = 0.44$, $p = .65$; total errors $F(2, 62) = 2.33$, $p = .11$].

Similarly, symptom validity scales were investigated for group differences for both child and parent report in the community recruited groups (see Table 7). With respect to child and adolescent self-report, multivariate analysis of BASC-2 validity scales indicated that the overall model was not significant [Wilk's $\Lambda = .78$, $F(8, 118) = 1.94$, $p = .06$]. However, examination of the individual validity scales indicated that simulators had higher scores on the V scale than control or clinical participants [$F(2, 62) = 5.71$, $p < .01$, partial $\eta^2 = .16$]. Children and adolescents did not differ on the responses that comprise the BASC-2 F Index [$F(2, 62) = 1.89$, $p = .16$], Consistency scale [$F(2, 62) = 0.87$, $p = .43$], or the L Index [$F(2, 62) = 1.03$, $p = .36$].

Table 7
*Community Recruited Child and Parent Symptom Validity Test MANOVA Results*

| | Control[a] n = 23 | Simulation[b] n = 20 | Clinical[c] n = 23 | *F* | Post hoc |
|---|---|---|---|---|---|
| BASC-2 Child | | | | | |
| F Index | .41 (.73) | 1.20 (2.07) | .65 (.94) | 1.89 | |
| Consistency | 7.50 (4.35) | 8.85 (5.90) | 9.30 (3.93) | 0.87 | |
| L Index | 3.18 (2.79) | 4.10 (3.09) | 4.52 (3.59) | 1.03 | |
| V Index | .14 (.47) | 1.70 (2.92) | .22 (.60) | 5.71** | (a = c) < b |
| BASC-2 Parent | | | | | |
| F Index | .13 (.34) | 2.60 (3.27) | .74 (1.48) | 8.64*** | (a = c) < b |
| Consistency | 7.22 (3.53) | 8.15 (2.94) | 8.65 (4.14) | 0.94 | |
| BRIEF Parent | | | | | |
| Negativity | .09 (.29) | 3.20 (2.88) | 1.61 (1.73) | 14.56*** | a < c < b |
| Inconsistency | 2.30 (1.64) | 2.60 (1.39) | 3.30 (1.49) | 2.63 | |

*Note:* *$p < .05$; **$p < .01$; ***$p < .001$; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function

Multivariate analysis of parent report BASC-2 validity scales indicated that groups differed [Wilk's $\Lambda = .76$, $F(4, 124) = 4.46$, $p < .01$]. Parents assigned to the simulation condition had higher scores on the F Index than parents in the clinical or control conditions [$F(2, 63) = 8.64$, $p < .001$, partial $\eta^2 = .22$]. This was still significant after utilization of a conservative alpha level due to violation of homogeneity of variance. Parents from control, clinical, and simulation groups exhibited consistency in their responses on the BASC-2 [$F(2, 63) = 0.94$, $p = .40$].

Parents also completed the BRIEF and the overall model indicated that groups differed [Wilk's $\Lambda = .64$, $F(4, 124) = 7.91$, $p < .001$]. Parents in the simulation condition had higher negativity scale scores than parents in control or clinical conditions [$F(2, 63) = 14.56$, $p < .001$, partial $\eta^2 = .33$] even with accounting for a more conservative alpha level given violation of Levene's test. Parents in all groups provided consistent BRIEF responses [$F(2, 63) = 2.63$, $p = .08$].

**Receiver Operating Curve Analyses**

In accordance with Aim 2, analyses were conducted to identify optimal cut-off scores for PVTs and SVTs. Initially, Receiver Operating Characteristic (ROC) curves were constructed by comparing the simulation group sequentially to the control and two clinical groups. Area Under the Curve (AUC) values were evaluated for acceptability for each PVT (see Table 8) and SVT (see Table 9). General guidelines for interpretation of the magnitude of discrimination of the AUC involve the following a) ≥ .90 outstanding, b) .80 to .90 excellent, c) .70 to .80 acceptable d) .60 to .70 fair, and e) .50 no discrimination (Hosmer & Lemeshow, 2000). Not surprisingly, discrimination of PVTs between the control and simulation groups was generally better than discrimination between simulation and clinical groups, as evidenced by higher ROC AUCs.

*Table 8*
PVT Area Under the Curve (AUC) Control and Clinical Groups Compared to the Simulation Group

| | Control | Community Clinical | MCW Clinical |
|---|---|---|---|
| | *n* = 23 | *n* = 23 | *n* = 125 |
| VSVT | | | |
| Total Easy Correct | **.76** | **.70** | .67 |
| Total Difficult Correct | **.75** | .61 | .66 |
| Total Items Correct | **.75** | .65 | .66 |
| Easy Latency[1] | .51 | .56 | .44 |
| Difficult Latency[1] | .59 | .55 | .51 |
| Digit Span | | | |
| DS ACSS | **.81** | **.73** | .65 |
| RDS | **.76** | **.71** | **.71** |
| RDS-R | **.79** | .69 | **.70** |
| WRAML-2 | | | |
| Response Bias | .51 | .60 | .53 |
| Discriminability % | .69 | .69 | .68 |
| d Prime | .68 | .70 | .69 |
| Forced Choice % Correct | **.74** | **.74** | -- |
| Rey FIT | | | |
| Recall Correct | .61 | .50 | -- |
| Recognition Correct | .65 | .64 | -- |
| False Positives | .66 | .63 | -- |
| Combination Score | .64 | .57 | -- |
| AST | | | |
| Alphabet Time[1] | **.79** | **.75** | -- |
| Alphabet Errors | .69 | .60 | -- |
| Counting Time[1] | **.85** | **.78** | -- |
| Counting Errors | .61 | .58 | -- |
| Days of Week Time[1] | **.82** | **.76** | -- |
| Days of Week Errors | .57 | .47 | -- |
| Months Time[1] | **.76** | .56 | -- |
| Months Errors | .62 | .45 | -- |
| Total Time[1] | **.81** | .68 | -- |
| Total Errors | .63 | .41 | -- |

*Note:* AUC = area under the curve; AUC values with at least acceptable discrimination are bolded (≥ .70); VSVT = Victoria Symptom Validity Test; Dif. = Difficult; [1] = time in seconds; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task

Table 9
Symptom Validity AUC for the Control and Community-Clinical Groups

| | Control | Community Clinical |
|---|---|---|
| BASC-2 Child | | |
| F Index | .58 | .52 |
| Consistency | .56 | .44 |
| L Index | .59 | .48 |
| V Index | .64 | .63 |
| BASC-2 Parent | | |
| F Index | **.81** | **.72** |
| Consistency | .56 | .49 |
| BRIEF Parent | | |
| Negativity | **.86** | .66 |
| Inconsistency | .56 | .37 |

*Note:* AUC = area under the curve; AUC values with at least acceptable discrimination are bolded (≥ .70); BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function

While it may be defensible to only examine classification statistics for PVTs and SVTs with at least acceptable AUC values, other literature highlights a drawback to this approach. For example, the AUC value summarizes test performance over the entire region of the ROC including areas in which attention would not be given (Lobo, Jiménez-Valverde, & Real, 2007). Thus, further investigation of classification statistics was conducted if multivariate analyses indicated statistically significant group differences and AUCs were at least fair. Given these requirements, VSVT latency scores, WRAML-2 response bias, Rey FIT recall and combination scores, and AST error scores were not investigated further.

## Classification Statistics and Cut-off Scores

Given, the similarities between the community- and medical center-clinical groups in demographics (see Table 1), memory (see Table 2), performance validity (see Table 6), and particularly in AUC values (see Table 7), those two groups were combined

for determination of diagnostic classification and optimal cut-off scores. While the medical-center clinical group and community-recruited clinical group differed in estimates of intellectual functioning and working memory (see Table 2), both groups were still within the average range. Further, the combined clinical group is primarily comprised of medical-center clinical participants who generally exhibit a higher degree of cognitive impairment than the community clinical groups. Thus, combining the groups leads to establishing more conservative PVT and SVT cut-off scores, which decreases the probability of incorrectly identifying a child or adolescent as putting forth insufficient effort when they are in fact trying to perform to the best of their ability.

For clarification, diagnostic classifications refer to sensitivity, specificity, positive predictive power, and negative predictive power. Sensitivity (Sn) reflects the proportion of individuals with the condition of interest (COI) that are correctly classified by the test (Berry & Schipper, 2008; Bianchini et al., 2001; Slick, 2006). Specificity (Sp) is the proportion of individuals without the COI that are correctly classified. Sensitivity and specificity can be combined into an index of test accuracy that specifies the odds or likelihood of positive or negative test results. Thus, a positive likelihood ratio (LR+) indicates the odds of a positive test result coming from a COI+ individual, whereas, a negative likelihood ratio (LR-) indicates the odds that a negative result came from a COI- individual. Positive predictive power (PPP) provides the probability that an individual with a positive test result has the COI, whereas, negative predictive power (NPP) is the probability that individuals with a negative test result do not possess the COI. These diagnostic classification statistics are also referred to as utility estimates and are utilized to develop the cut-off scores, or benchmarks, to denote passing or failure of a SVT or

PVT (Strauss et al., 2006). Cut-off scores are considered optimal when specificity is at least 90% and sensitivity is maximized.

Complete classification statistics (e.g., Sn, Sp, LR+, LR-, PPP, NPP) for every possible cut-off score for each PVT and SVT are presented in Appendices J through U. Only the optimal cut-off scores for PVTs and SVTs are presented in the following tables and text. Selection of optimal cut-off scores involved specificity of at least .90 while maximizing sensitivity.

In general, PVT cut-off scores (see Table 10) were more conservative for clinical participants than control participants. While it was hypothesized that stand-alone measures of performance (see VSVT, Rey FIT, AST) would display better classification statistics than embedded measures (see embedded Digit Span, WRAML-2), that was not always the case. Optimal cut-off scores for most PVTs displayed moderate sensitivity when specificity was maximized. However, the Rey FIT and the AST Months of the Year condition displayed weak sensitivity. WRAML-2 discriminability, d prime, and the forced choice task show promise as novel embedded PVTs given that they display classification statistics that are consistent with, or better than, previously identified PVTs.

Table 10
*Optimal PVT Cut-off Scores and Classification Statistics*

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score Control/ Clinical | Cumulative % of sample at cut-score Simulation |
|---|---|---|---|---|---|---|---|---|---|
| VSVT | | | | | | | | | |
| Easy Correct | | | | | | | | | |
| Control | ≤ 21 | .47 | .95 | 10.42 | .55 | .90 | .68 | 4.55% | 47.37% |
| Clinical | ≤ 17 | .37 | .93 | 5.45 | .68 | .41 | .92 | 6.76% | 36.84% |
| Dif. Correct | | | | | | | | | |
| Control | ≤ 18 | .53 | .95 | 11.58 | .50 | .91 | .70 | 4.54% | 52.63% |
| Clinical | ≤ 12 | .32 | .91 | 3.60 | .75 | .32 | .91 | 8.78% | 31.58% |
| Total Correct | | | | | | | | | |
| Control | ≤ 34 | .47 | .95 | 10.42 | .55 | .90 | .68 | 4.55% | 47.37% |
| Clinical | ≤ 31 | .42 | .90 | 4.15 | .64 | .35 | .92 | 10.14% | 42.11% |
| Digit Span | | | | | | | | | |
| DS ACSS | | | | | | | | | |
| Control | ≤ 7 | .55 | .95 | 12.10 | .47 | .92 | .70 | 4.55% | 55.00% |
| Clinical | ≤ 4 | .50 | .92 | 6.59 | .54 | .48 | .93 | 7.59% | 50.00% |
| RDS | | | | | | | | | |
| Control | ≤ 6 | .50 | .96 | 11.50 | .52 | .91 | .69 | 4.35% | 50.00% |
| Clinical | ≤ 5 | .45 | .92 | 5.81 | .60 | .45 | .92 | 7.75% | 45.00% |
| RDS-R | | | | | | | | | |
| Control | ≤ 10 | .50 | .91 | 5.50 | .55 | .83 | .67 | 9.09% | 50.00% |
| Clinical | ≤ 8 | .50 | .92 | 6.15 | .54 | .50 | .92 | 8.13% | 50.00% |
| WRAML-2 | | | | | | | | | |
| Discriminability | | | | | | | | | |
| Control | ≤ 87.50 | .60 | .91 | 6.60 | .44 | .86 | .71 | 9.09% | 60.00% |
| Clinical | ≤ 75.00 | .55 | .95 | 11.14 | .47 | .73 | .90 | 6.00% | 55.00% |
| d Prime | | | | | | | | | |
| Control | ≤ 3.0 | .60 | .91 | 6.60 | .44 | .86 | .71 | 4.50% | 60.00% |
| Clinical | ≤ 1.5 | .55 | .93 | 7.37 | .49 | .69 | .87 | 7.50% | 55.00% |
| Forced Choice % | | | | | | | | | |
| Control | ≤ 92 | .50 | .95 | 11.00 | .52 | .91 | .68 | 4.55% | 50.00% |
| Clinical | ≤ 92 | .50 | 1.00 | 0.00 | .50 | 1.00 | .70 | 0.00% | 50.00% |
| Rey FIT | | | | | | | | | |
| Recognition | | | | | | | | | |
| Control | ≤ 6 | .35 | 1.00 | 0 | .65 | 1.00 | .63 | 0.00% | 35.00% |
| Clinical | ≤ 3 | .20 | .91 | 2.30 | .88 | .67 | .57 | 8.70% | 20.00% |
| False Positives | | | | | | | | | |
| Control | ≥ 2 | .30 | .91 | 3.30 | .77 | .75 | .59 | 4.55% | 30.00% |
| Clinical | ≥ 2 | .30 | .96 | 6.90 | .73 | .86 | .61 | 4.35% | 30.00% |
| AST | | | | | | | | | |
| Alphabet | | | | | | | | | |
| Control | ≥ 14 | .55 | .91 | 6.05 | .50 | .85 | .69 | 9.09% | 55.00% |
| Clinical | ≥ 19 | .40 | .91 | 4.60 | .66 | .80 | .64 | 8.70% | 40.00% |
| Count to 20 | | | | | | | | | |
| Control | ≥ 7 | .65 | .95 | 14.3 | .37 | .93 | .75 | 4.55% | 65.00% |
| Clinical | ≥ 8 | .55 | .91 | 6.33 | .49 | .85 | .70 | 8.70% | 55.00% |
| Days of Week | | | | | | | | | |
| Control | ≥ 5 | .50 | 1.00 | 0 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| Clinical | ≥ 6 | .45 | .96 | 10.35 | .58 | .90 | .67 | 4.35% | 45.00% |

Table 10 continued on next page

Table 10
*Optimal PVT Cut-off Scores and Classification Statistics*

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score Control/ Clinical | Simulation |
|---|---|---|---|---|---|---|---|---|---|
| AST | | | | | | | | | |
| Months of Year | | | | | | | | | |
| Control | ≥ 23 | .10 | .91 | 1.1 | .99 | .50 | .53 | 9.09% | 10.00% |
| Clinical | ≥ 22 | .15 | .91 | 1.73 | .93 | .60 | .55 | 8.70% | 15.00% |
| Total Time | | | | | | | | | |
| Control | ≥ 45 | .50 | .91 | 5.50 | .55 | .83 | .67 | 9.09% | 50.00% |
| Clinical | ≥ 52 | .40 | .91 | 4.60 | .66 | .80 | .64 | 8.70% | 40.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; VSVT = Victoria Symptom Validity Test; Dif. = Difficult; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task

With respect to SVT cut-off scores, one scale from each questionnaire (e.g., BASC-2 child report V Index, BASC-2 parent report F Index, and BRIEF parent report Negativity scale) was identified as distinguishing simulation participants from control or clinical participants (see Table 11). SVT sensitivity was slightly weaker than PVTs when specificity was maximized.

Table 11
*Community Recruited Control and Clinical SVT Classification Statistics*

| SVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Control | Simulation |
| BASC-2 Child V Index | | | | | | | | | |
| Control | ≥ 2 | .35 | .95 | 7.70 | .68 | .88 | .62 | 4.55% | 35.00% |
| Clinical | ≥ 2 | .35 | .91 | 4.03 | .71 | .78 | .62 | 8.70% | 35.00% |
| BASC-2 Parent F Index | | | | | | | | | |
| Control | ≥ 3 | .40 | 1.00 | 0 | .60 | 1.00 | .66 | 0.00% | 40.00% |
| Clinical | ≥ 4 | .35 | .96 | 8.05 | .68 | .88 | .63 | 4.35% | 35.00% |
| BRIEF – Parent Negativity | | | | | | | | | |
| Control | ≥ 1 | .75 | .91 | 8.63 | .27 | .88 | .81 | 8.70% | 75.00% |
| Clinical | ≥ 5 | .35 | .96 | 8.05 | .68 | .88 | .63 | 4.35% | 35.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function

## Correlations between PVTs and SVTs

Consistent with Aim 3, associations between PVTs and SVTs for the measures that effectively distinguished simulation participants from clinical and control participants were explored. Associations varied across measures (see Table 12). The BASC-2 parent report F Index was moderately, negatively associated with child performance on the VSVT difficult items ($r(43) = -.30$, $p < .05$), embedded Digit Span measures (DS ACSS $r(43) = -.32$, $p < .05$; RDS $r(43) = -.32$, $p < .05$; RDS-R $r(43) = -.46$, $p < .01$), and Rey FIT recognition ($r(43) = -.34$, $p < .05$). The BASC-2 parent report F Index was moderately, positively associated with their child's Rey FIT false positive score ($r(43) = .39$, $p < .01$). The BASC-2 self-report V Index was negatively, moderately correlated with WRAML-2 d prime ($r(43) = -.34$, $p < .05$) and Rey FIT recognition performance ($r(43) = -.32$, $p < .05$), while the BASC-2 self-report V Index was positively

associated with AST Counting ($r(43) = .45$, $p < .01$) and Total Time ($r(43) = .52$, $p < .01$).

Table 12
Correlations between PVTs and SVTs for Community Recruited Control and Clinical Participants

| | Brief Negativity | BASC-2 Parent F Index | BASC-2 Child V Index |
|---|---|---|---|
| VSVT | | | |
| Easy Correct | .03 | -.16 | -.22 |
| Difficult Correct | -.18 | -.30* | -.27 |
| Total Correct | -.13 | -.28 | -.27 |
| Digit Span | | | |
| DS ACSS | -.10 | -.32* | -.14 |
| RDS | -.01 | -.32* | -.16 |
| RDS-R | -.07 | -.46** | -.24 |
| WRAML 2 | | | |
| Discriminability % | .06 | -.03 | -.25 |
| d Prime | .14 | -.05 | -.34* |
| Forced Choice % | .09 | .06 | .05 |
| Rey FIT | | | |
| Recognition | -.08 | -.34* | -.32* |
| False Positives | .20 | .39** | .02 |
| AST | | | |
| Alphabet | -.01 | -.09 | .16 |
| Counting | -.09 | .23 | .45** |
| Days of Week | -.16 | -.05 | .32* |
| Months | .13 | .20 | .28 |
| Total | .03 | .13 | .52** |

*Note:* *$p < .05$; **$p < .01$; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function; VSVT = Victoria Symptom Validity Test; Dif. = Difficult; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task

Associations between PVTs and SVTs for simulation participants were conducted separately from the control and clinical participants, given inherent differences in instructions (see Table 13). Parent-report SVTs, BRIEF Negativity and BASC-2 F Indices, were not meaningfully associated with any youth PVT performances. However, the child and adolescent BASC-2 V Index was significantly, moderately associated with nearly all PVT measures, except the Rey FIT false positives and AST counting and alphabet completion time.

Table 13
*Correlations between PVTs and SVTs for Community Recruited Simulation Participants*

| | Brief Negativity | BASC-2 Parent F Index | BASC-2 Child V Index |
|---|---|---|---|
| VSVT | | | |
| Easy Correct | .12 | .09 | -.46* |
| Difficult Correct | .20 | .15 | -.49* |
| Total Correct | .17 | .13 | -.51* |
| Digit Span | | | |
| DS ACSS | .12 | -.07 | -.50* |
| RDS | .28 | -.02 | -.48* |
| RDS-R | .16 | -.11 | -.51* |
| WRAML 2 | | | |
| Discriminability % | .13 | .01 | -.50* |
| d Prime | .12 | -.01 | -.49* |
| Forced Choice % | .27 | -.06 | -.59** |
| Rey FIT | | | |
| Recognition | .02 | -.14 | -.71** |
| False Positives | .01 | .08 | .21 |
| AST | | | |
| Alphabet | -.20 | -.04 | .44 |
| Counting | -.31 | -.03 | .40 |
| Days of Week | -.09 | -.02 | .46* |
| Months | -.12 | -.16 | .60* |
| Total | -.24 | -.05 | .50* |

*Note:* *$p < .05$; **$p < .01$; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function; VSVT = Victoria Symptom Validity Test; Dif. = Difficult; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised; WRAML-2 = Wide Range Assessment of Memory and Learning; Rey FIT = Rey 15-Item Test; AST = Automatized Sequences Task

## Discussion

Neuropsychological assessment relies on valid self-report and credible performance. Symptom exaggeration and misrepresentation of abilities may confound interpretation of neuropsychological test data, subsequent diagnosis, and treatment. The development of PVTs and SVTs has provided psychometric tools to evaluate the validity of test performance and self- and parent-report of symptoms. While PVTs and SVTs have been extensively studied in adult populations, literature regarding measures validated in child and adolescent samples is still emerging (DeRight & Carone, 2015; Kirkwood, 2015).

Developmental research suggests that children are capable of deception (Talwar & Lee, 2002; Talwar et al., 2007) and that sophistication of deception improves with development of executive functioning (see Anderson, 2002; Walczyk et al., 2003) through adolescence (Salekin et al., 2008). While deception can be volitional, children and adolescents may also exhibit invalid profiles due to presence of a conversion or factitious disorder, or due to more nuanced or unknown reasons (see Flaro et al., 2007). Thus, it is not surprising that invalid performance and symptom report occur in child and adolescent clinical, forensic, and educational evaluations. Estimated base rates of invalid performance in clinical contexts range from 2 to 20% (see Kirkwood, 2015) and are much higher in disability contexts (26 to 60%; Chafetz, 2008; Chafetz et al., 2007). Rates of invalid performance and symptom report in child and adolescent psychoeducational evaluations are currently unknown; however, case studies document response distortion (Harrison et al., 2012; Lu & Boone, 2002). Clearly, there is a need to systematically examine how children and adolescents engage in response distortion during psychoeducational evaluations.

While evidence of invalid performance and symptom report in child and adolescent evaluation is available in the literature and highlights the need for PVT and SVT use, consensus statements by prominent neuropsychological organizations also encourage pediatric validity assessment research (AACN Board of Directors, 2007; Bush et al., 2005; Heilbronner et al., 2009). Further, discussion of ethical considerations in child and adolescent assessment highlights the need for systematic evaluation of performance and symptom validity to prevent erroneous conclusions regarding functioning and subsequent treatment (MacAllister & Vasserman, 2015). Notably, a

recent survey documents a majority of pediatric neuropsychologists incorporate at least one PVT (92%) or at least one SVT (88%) within assessments (Brooks, Ploetz, & Kirkwood, 2016). Thus, it appears that routine PVT and SVT use is becoming an accepted practice.

A primary challenge within child and adolescent PVT and SVT literature relates to available cut-off scores and the respective samples that scores and classification statistics are derived from. For example, survey results indicate that pediatric neuropsychologists most commonly utilize embedded PVTs, specifically; RDS, CVLT-C discriminability index, and CVLT-II forced choice (Brooks et al., 2016). However, recommended cut-off scores for RDS and the CVLT-C discriminability index vary depending on sample (e.g., mTBI see Baker et al., 2014; Kirkwood et al., 2011; or neurological conditions see Brooks & Ploetz, 2015; Perna et al., 2014). Further, the CVLT-II Forced Choice cut-off score has not been validated with adolescents (Schwartz et al., 2016). Frequently in child and adolescent PVT literature, studies have described rates at which youth can pass PVTs at adult derived cut-off scores (e.g., Brooks, 2012), have utilized youth control and simulation groups without corresponding clinical groups (e.g., Blaskewitz et al., 2008), or commonly utilized cut-off scores are derived from mTBI samples which are inherently less cognitively impaired than other medical, neurological, or developmental populations (e.g., Kirkwood et al., 2014). Thus, interpretation of some available PVT cut-off scores can be challenging. Further, a paucity of research related to parent and child SVTs outside of standardization samples exists, even though neuropsychologists report they commonly utilize the BRIEF and BASC-2 SVTs (Brooks et al., 2016).

Further, some clinicians have expressed concerns regarding the use of PVTs and SVTs with young children or individuals with cognitive impairment (Brooks et al., 2016). These concerns are consistent with some literature that has highlighted limitations of PVTs with young children (e.g., Rey FIT see Constantinou & McCaffrey, 2003, or Lichtenstein et al., 2017; Computerized Assessment of Response Bias (CARB); Allen, Conder, Green, & Cox, 1997, see Courtney et al., 2003; Word Memory Test WMT; Green, 2003 see Courtney et al., 2003). Research that documents PVT performance from children and adolescents with cognitive disabilities (Carone, 2014; Gidley-Larson et al., 2015; Green & Flaro, 2014) is still emerging. Thus, concerns related to the use of PVTs and SVTs appear to have merit, given the current body of literature. Thus, the present study sought to evaluate those concerns in cut-off score validation.

The primary goal of the present study was to identify youth appropriate validity cut-off scores for children and adolescents with various neurological, medical, and developmental conditions. Therefore, through a recommended validation approach utilizing control, simulation, and two clinical groups, the present study sought to investigate numerous PVTs and parent- and child-report SVTs. Non-clinical participants were assigned to control or simulation conditions using a stratified method based upon on age and sex. The two clinical groups in the present study were intentionally selected; the medical-center pediatric neuropsychological practice represents standard clinical practice, and a community-recruited clinical group eliminates potential secondary gain issues associated with securing a diagnosis.

An important component of this project was to evaluate whether children, adolescents, and parents can simulate. In response to a brain injury scenario, children and

adolescents did simulate cognitive impairment on performance tasks (WRAML-2 Verbal Learning, WISC-V Digit Span) by performing at a level below clinical participants. This finding is consistent with previous literature that has documented that children and adolescents can feign cognitive impairment (see Blaskewitz et al., 2008; Gunn et al., 2010; Lucio et al.,, 2002; McKinzey et al., 2003; Rambo et al., 2015). With respect to SVT completion, while youth altered response patterns on a self-report questionnaire (BASC-2) to elevate one validity scale, clinical scales reflecting psychological issues were not elevated. In a novel aspect of this research, parents altered their response patterns on two observer-report questionnaires (BRIEF and BASC-2) to a degree that generally reflected clinical impairment in emotional and behavioral domains. Therefore, our hypothesis was partially supported. It appears that youth could suppress cognitive performance, but they may not have possessed insight into how feigned cognitive symptoms might affect their emotional or behavioral functioning or the ability to complete daily activities. Whereas, parents could exaggerate symptoms indicating that they may have been able to consider how cognitive symptoms might affect emotional, behavioral, and adaptive functioning.

With respect to PVT performance in the current study, cut-off scores from previous studies were considered to determine false positive rates. Additionally, PVTs were evaluated for relationships with age and cognitive tasks. Finally, cut-off scores derived from the present study are presented and discussed. These considerations are presented for all PVTs in the study in the following order: VSVT, embedded Digit Span measures, embedded WRAML-2 measures, Rey FIT, and AST.

**Victoria Symptom Validity Test**

Investigation of currently available VSVT cut-off scores derived from adult standardization samples in the manual revealed that the majority of control, community-clinical, and medical center-clinical participants could 'pass' VSVT easy items (≥ 16 easy items correct, ≥ 95% passed) and total items (≥ 30 total items correct ; ≥ 91% passed). However, only 78% of the community-clinical and 81% of the medical-center clinical participants could achieve a score in the valid range for VSVT difficult items correct. The only other study to investigate the VSVT in youth sample described similar pass rates (Brooks, 2012). Further, only 76% of children younger than 10 years of age and only 43% of individuals with an intellectual disability could achieve a VSVT difficult items score in the manual recommended valid range. It is clear that it would be inappropriate to apply adult-derived cut-off scores to a youth clinical sample.

Thus, empirically derived VSVT cut-off scores are necessary for youth with clinical conditions. Optimal cut-off scores were proposed based upon achievement of at least 90% specificity while maximizing sensitivity. Based upon clinical participant performance, optimal cut-off scores to indicate invalid performance are: VSVT easy items ≤ 17 (Sn = .37, Sp = .93), difficult items ≤ 12/13 (Sn = .32, Sp = .91), and total items ≤ 31 (Sn = .42, Sp = .90); however, a clinician may wish to further maximize specificity to avoid false positives by utilizing other scores proposed in Appendix J (e.g., VSVT easy items ≤ 7 Sn = 0.00, Sp = 1.00), though sensitivity is sacrificed. Other scores investigated such as VSVT item response latencies were not useful for discrimination of simulation participants from control or clinical participants.

Notably, VSVT performance was significantly associated with age, estimates of intellectual ability, and working memory, but not verbal memory. Similarly, Brooks (2012) found VSVT performance was related to age, intelligence, processing speed, but not sustained attention or memory. Therefore, clinicians may wish to consider this information when selecting a PVT to utilize if there are known cognitive deficits. Nonetheless, VSVT cut-off scores presented above are considered appropriate for a youth clinical sample.

**Embedded Digit Span Performance Validity Measures**

Regarding embedded digit span measures, RDS (≤ 6) and ACSS (≤ 5) cut off scores have been previously established in an mTBI sample (Kirkwood et al., 2011), and are higher than those derived utilizing more heterogeneous clinical samples (≤ 4; Loughan et al., 2012; Perna et al., 2014). The cut-off score established with a mTBI sample resulted in a high degree of false positives (35-39%) in samples of children and adolescents with epilepsy (Welsh et al., 2012) and a non-clinical community sample (Blaskewitz et al., 2008). Similarly, the present study documents that if cut-off scores derived from an mTBI sample are applied, there is an unacceptable number of false positives within this mixed clinical sample (RDS ≤ 6 =17-20%; DS ACSS ≤ 5 = 13 – 26%). Understandably, false positive risk was even higher with children younger than 10 (15-24%) and youth with intellectual disability (50-64%). Proposed cut-off scores in this mixed clinical sample displayed moderate sensitivity when specificity was optimized (DS ACSS ≤ 4 Sn = .50, Sp = .92; RDS ≤ 5 Sn = .45, Sp = .92). Though an RDS ≤ 4 (Sn = .40, Sp = .99) cut-off score in this mixed clinical sample was consistent with

classification statistics presented elsewhere (Loughan et al., 2012; Perna et al., 2014), there is a decrease in sensitivity.

While RDS has been investigated with children and adolescents (e.g. Araujo et al., 2014; Blaskewitz et al., 2008; Kirkwood, et al., 2011; Welsh et al., 2012), the utility of incorporating reliably accurate performance on the sequencing trial had not yet been considered as an indicator of task engagement. RDS-R has exhibited more optimal classification statistics than RDS in preliminary adult studies (e.g. Reese et al., 2012; Spencer et al., 2013; Young et al., 2012). The proposed adult RDS-R cut-off score (≤ 11) was clearly not appropriate for a youth sample due to high false positive rates in clinical participants (39%), children younger than 10 (32%), and youth with intellectual disability (57%). A novel RDS-R proposed cut-off score (≤ 8, Sn = .50, Sp = .92) displayed similar sensitivity and specificity to other embedded digit span measures (i.e., RDS; DS ACSS). Nevertheless, clinicians and researchers may wish to utilize this embedded measure because it incorporates the entire WISC-V digit span task for a slightly larger and more continuous evaluation of task engagement.

It is important to keep in mind that RDS and RDS-R are derived from a task that assesses simple verbal attention and working memory (Wechsler, 2014), which can be affected by multiple neurological conditions. Supporting the notion that Digit Span performance is associated with cognitive constructs, RDS and RDS-R were associated with estimates of intellectual functioning and verbal memory. It is reasonable that RDS and RDS-R are correlated with other areas of cognitive functioning given that these scores are raw scores (i.e., not age corrected) and cognitive functions are generally related. Importantly, a recent survey of pediatric neuropsychologists reported that RDS is

the *most* commonly utilized measure of child and adolescent performance validity (Brooks et al., 2016). While survey respondents did not indicate what cut-off score is most commonly applied, the current findings suggest that some published RDS cut-off scores are problematic. Consideration of the cut-off scores proposed in the present study is warranted and will likely decrease the probability of false positive errors in clinical decision making.

**Embedded WRAML-2 Performance Validity Measures**

In adult PVT literature, a parametric signal detection statistic, d prime, and non-parametric signal detection statistics, recognition discriminability and response bias, have been effectively utilized as embedded PVTs on list learning tasks (Delis et al., 2000, see Curtis et al., 2006; Wolfe et al., 2010). These statistics were derived from WRAML-2 Verbal Learning Recognition task performance. Response bias was not useful in discrimination of simulation participants from control or clinical participants. Discriminability (≤ 75.00 Sn = .55, Sp = .95) and d prime (≤ 1.5 Sn = .55, Sp = .93) cut-off scores displayed moderate sensitivity with optimized specificity in this mixed clinical sample. Presently, discriminability and d prime WRAML-2 Verbal Learning statistics are not included in the WRAML-2 scoring program, so clinicians and researchers would need to generate them, but initial evidence suggests that they are a useful embedded tool.

Related to the previously described embedded WRAML-2 PVTs, some researchers have investigated similar measures in the CVLT-C with youth. For example, in a mTBI sample, an age-corrected z-score of -0.5 from the CVLT-C Recognition Discriminability index optimally identified individuals with invalid performance (Baker

et al., 2014); whereas, in a mixed clinical group, a much more extreme age-corrected z-score of -3.0 was optimal (Brooks & Ploetz, 2015). These vastly different cut scores clearly reflect underlying sample characteristics, meaning that the mTBI sample is inherently less impaired, whereas, individuals with varied neurological conditions will demonstrate bona-fide memory impairments. This obviously presents a challenge for a clinician, given that testing determines the level of functioning so one cannot necessarily select a PVT cut-off score a-priori. Certainly the goal is to optimize specificity and sensitivity, though it is generally thought to be more important to limit false positives (i.e., maximize specificity). Similar to Brooks and Ploetz (2015), the present study included a mixed clinical sample with varied medical, neurological, and developmental conditions in which memory and learning may be affected. Thus, the cut-off scores derived for the WRAML-2 embedded measures are likely to exhibit poorer sensitivity in a less cognitively impaired sample.

Additionally, a novel WRAML-2 Forced Choice recognition task was created as an embedded PVT, which displayed moderate sensitivity when specificity was optimized (Forced choice percent correct $\leq$ 92%, Sn = .50, Sp = 1.00). Dependent on age, children or adolescents are administered a different number of Forced Choice items due to the number of initial words on the learning trials, thus, children 8 years and younger must identify 12 out of 13 items correctly on the forced choice task. Whereas, children 9 years and older must identify 15 out of 16 items correctly. Encouragingly, the WRAML-2 forced choice task was not associated with age, estimates of intellectual functioning, working memory, or verbal memory. Therefore, the forced choice task exhibits some benefit over the other WRAML-2 embedded measures. In a similar study, Lichtenstein

and colleagues (2017) developed a forced choice task for the CVLT-C (FCR-C raw score ≤ 13, Sn = .15, Sp = .94), which displayed poor sensitivity when specificity was optimized. The WRAML-2 forced choice task performed slightly better in terms of sensitivity in our sample than the CVLT-C forced choice task.

**Rey Fifteen Item Test**

Previous Rey FIT research with non-clinical youth suggested that a Rey FIT recall cut-off score of ≤ 7 was optimal (Blaskewitz et al., 2008), whereas a slightly higher cut-off score (< 9) was optimal in youth with a mTBI (Green et al., 2014). In the current community recruited sample, applying a Rey FIT recall cut-off score from a non-clinical sample (≤ 7) resulted in approximately 13% of the clinical group being identified as providing invalid performance, further all children scoring below that cut-off score were younger than 10 years of age. The latter finding is consistent with previous research indicating an association between Rey FIT performance and age, and children under 10 years have displayed higher rates of failure (Constantinou &McCaffrey, 2003; Lichtenstein et al., 2017).

The Rey FIT Combination Score (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002) takes into account the total number of correctly recalled items, correctly recognized items, and recognition error rate. A Rey FIT combination score derived from a youth mTBI sample (< 26; Green et al., 2014) would have identified large percentages of children and adolescents across control (41%), clinical (35%), children under 10 years old (52%), and youth with intellectual disability (50%) groups as providing invalid performance. Thus, the youth Rey FIT combination score cut-off score proposed by

Green and colleagues (2014) appears to be problematic, as well, and associated with unacceptable false positive rates.

All possible Rey FIT scores were investigated and the recall and combination scores described above did not differentiate simulation participants from control or clinical participants. Whereas, the Rey FIT Recognition total correct score (≤ 3 Sn = .20, Sp = .91) and recognition False Positive score (≥ 2 Sn = .30. Sp = .96) differentiated simulators from other participants. However, sensitivity was the lowest for these measures compared to other PVTs investigated in the present study. While stand-alone PVTs often perform more optimally than embedded PVTs, this does not appear to be the case when comparing the Rey FIT to embedded Digit Span and WRAML-2 PVTs.

**Automatized Sequences Task**

The AST was first developed as a PVT for a youth mTBI clinical sample (Kirkwood et al., 2014a). Application of the cut-off scores derived from the mTBI sample would have indicated higher than acceptable rates of invalidity across select groups (control 14-23%, clinical 22-48%, and children under 10 years of age 22-41%). Though, notably, children with intellectual disability passed the AST conditions at the cut-off scores proposed by Kirkwood and colleagues (2014). Optimal cut-off scores for the mixed clinical group for each condition in seconds are as follows: alphabet ≥ 19 (Sn = .40, Sp = .91), counting to 20 ≥ 8 (Sn = .55, Sp = .91), days of the week ≥ 6 (Sn = .45, Sp = .96), months of the year ≥ 22 (Sn = .15, Sp = .91), and total time for the four conditions ≥ 52 (Sn = .40, Sp = .91). Of note, the months of the year task did not actually differentiate simulators from control or clinical participants. In fact, many children had

difficulty reciting the months of the year. Therefore, for purposes of exploration, a revised total time score for the three conditions (≥ 38, Sn = .40, Sp = .91) was generated. When specificity was optimized, sensitivity was slightly lower than the original AST total time. Novel error rates for each AST condition were explored in the present study, but they were not useful for differentiating simulators from control or clinical participants.

In summary, the hypothesis that stand-alone PVTs would exhibit better classification statistics than embedded measures (see Bianchini et al., 2001) was not supported in the current study. While the sensitivities for the stand-alone measures (e.g., VSVT, Rey FIT) were acceptable, multiple embedded measures were more optimal in differentiating simulators and a mixed clinical sample. The WRAML-2 signal detection measures (Discriminability, d Prime), WRAML-2 Verbal Learning Forced Choice, AST, and adjusted RDS, DS ACSS, and RDS-R show promise with sensitivities ranging from .40 to .55. Further, applying PVT cut-off scores derived from youth mTBI samples routinely resulted in high false positive rates in groups consisting of mixed clinical conditions, young children, and youth with intellectual disability. Clinicians should recognize an increased probability of incorrectly identifying invalid performance if mTBI derived cut-off scores are applied to other clinical populations.

**Symptom Validity Measures**

PVTs and SVTs are complimentary and may help clinicians make decisions regarding the validity of a patient's overall presentation. As mentioned previously, while SVTs have been extensively investigated in adult samples, literature regarding their

utility in youth neuropsychological evaluations is limited. Thus, the present study sought to investigate SVT utility. Further, previous research indicates that invalidity on a PVT does not necessarily equate with invalidity on SVTs (Kirk et al., 2014; VanDyke et al., 2013). PVTs and SVTs may assess different constructs, thus the present study sought to evaluate their convergence.

Regarding SVTs in the present study, child and adolescent report on the BASC-2 indicated that only the V Index, a scale designed to detect random responding, distinguished simulators from control or clinical participants, while other validity scales (e.g., F Index) were not beneficial. Classification statistics suggested that a V Index cut-off score of ≥ 2 was optimal, which is classified by the BASC-2 manual as "questionable" for interpretation. Of note, the BASC-2 manual suggests caution when interpreting self-report measures with a V Index raw score of ≥ 4. None of the individuals in the community-clinical or control group exhibited invalid performance on V Index at the manual suggested cut-off scores. The majority of individuals in the simulation group ($n$ = 15, 75%) also exhibited valid performance on the V Index. This finding is generally consistent with a broader literature.

Somewhat similarly, Kirk and colleagues (2014) reported that children and adolescents who sustained a mTBI and failed a PVT did not provide invalid response patterns on the BASC-2. However, Kirkwood and colleagues (2014b) found that children and adolescents in a mTBI sample who failed PVTs reported significantly more post-concussion symptoms on rating forms than youth who passed PVTs. Notably, none of the BASC-2 validity scales were designed to detect over-reporting of cognitive or somatic symptoms. Thus, it is logical that BASC-2 validity scales would be unlikely to detect

simulators feigning cognitive and physical symptoms associated with the neuropsychological scenario provided or similar real-world clinical contexts (e.g., medical/neurological contexts). Our hypothesis, that PVTs, relative to SVTs, would more effectively differentiate groups was supported in this context. While PVTs and SVTs can be complimentary, youth may pass an SVT and not exhibit valid neuropsychological performance. Conversely, youth that pass PVTs may provide invalid symptom report, since current SVTs and PVTs assess different constructs.

Other SVT research is limited and directed towards identifying psychopathology, not cognitive or somatic complaints. Previous adolescent SVT research conducted with the MMPI-A utilized scenarios in which simulators were asked to feign psychological disorders indicated that MMPI-A *F*-family of validity scales discriminated simulators from clinical and control participants (*F* see Baer et al., 2010; Stein et al., 1995; *F, F1, F2* see Lucio et al., 2002; *F-K* see Rogers et al., 1996). The MMPI-A *VRIN* scale also effectively discriminated random responding from clinical participants (Baer et al., 2010). Of note, in the present study, BASC-2 clinical scales were not elevated by simulators, possibly, suggesting that the brain injury scenario provided may not prompt individuals to over-report psychological distress. Further, the MMPI-A studies only included adolescents, whereas, the present study included younger children. While children and adolescents were asked to respond in a manner that might assist them in receiving accommodations and provided with common brain injury symptoms, they may not have possessed the insight to consistently alter response patterns on items that were less obviously related to the scenario or relevant to populations with bona-fide impairments.

In contrast to youth samples, SVT scales relevant to over-reporting of cognitive and somatic symptoms have been extensively investigated in adult populations. For example the Minnesota Multiphasic Personality Inventory – Second Edition (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) possess validity scales designed to detect over-reporting of somatic and cognitive symptoms. For example, the MMPI-2 Response Bias Scale meaningfully predicts PVT failure (Whitney, Davis, Shepard, & Herman, 2008). Additionally, multiple MMPI-2-RF validity scales have demonstrated strong associations with invalid PVT performance (Gervais, Wygant, Roger, Sellbom, & Ben-Porath, 2011) and malingered neurocognitive dysfunction (Tarescavage, Wygant, Gervais, & Ben-Porath, 2012). While other researchers investigating validity within youth mTBI populations have suggested that further development and investigation of child and adolescent SVTs within contexts relevant to over-reporting of cognitive or somatic symptoms is necessary (Kirk et al., 2014; Kirkwood et al., 2014b), this research has yet to be initiated.

A particularly novel component of the present study, involved the inclusion of the BRIEF and BASC-2 parent-report measures with group assignment matched to their child or adolescent. While recent pediatric neuropsychologist survey data indicates that the BRIEF and BASC-2 questionnaires are the most common SVTs utilized in practice (Brooks et al., 2016), no youth PVT or SVT validation studies conducted to date include investigation of parent-report. Parents may consciously or unconsciously experience a desire to for the child to receive academic accommodations, various types of treatment, or disability benefits (Chafetz & Dufrene, 2014; Chafetz & Prentkowski, 2011). Thus,

our goal was to investigate the ability of parents to feign symptom report and compare response patterns to parents of youth with and without clinical conditions.

The present study revealed that the BASC-2 F Index and the BRIEF Negativity scale effectively discriminated parents within the simulation condition from parents in the control or clinical conditions. Both scales are designed to detect overly negative appraisals of child or adolescent emotional and behavioral functioning. However, very few parent SVTs within the simulation group were elevated to the level of invalidity suggested by the administration manuals (e.g., BRIEF Negativity ≥ 9; *n* = 1, 5%; BASC-2 F Index ≥ 6, *n* = 2, 10%). In this context, the majority of parents were essentially able to avoid invalidity detection, without specific instructions to do so. Thus, in actual clinical or research contexts, these parents while intentionally simulating, they would not be detected by validity scales and clinical scales would simply appear elevated. This may lead a clinician or researcher to interpret the questionnaire as valid and utilize their responses to support a diagnosis or research findings.

Parents in the simulation condition also altered their responses to the degree that clinical scales were elevated. Each BASC-2 and BRIEF clinical scales from simulation parents were meaningfully greater than parents in clinical and non-clinical groups asked to appraise their child's typical emotional and behavioral functioning. Further, clinical scale averages for simulating parents were frequently elevated to the sub-clinical or clinical range. Thus, when parents were given a false brain injury scenario and asked to respond in a manner that would ensure academic accommodations or other supports, profiles were infrequently identified as invalid, and clinical scales were elevated to a degree that a clinician would likely interpret impairment. Similar to youth self-report,

current parent-report SVTs are not specifically designed to detect over-reporting of cognitive or somatic symptoms. Given frequent reliance on parent-report for diagnosis or treatment recommendation, findings from youth literature of over-reporting post-concussive symptoms (Kirkwood et al., 2014b), and adult cognitive and somatic over-reporting (Gervais et al., 2011; Tarescavage et al., 2012), additional investigation is certainly warranted. Further, development of a parent SVT scale more specific to cognitive and somatic over-reporting appears necessary. One method to develop BASC or BRIEF validity scales specific to feigned cognitive impairment might include investigating items responses from individuals who do and do not pass PVTs. Similar to the development of RBS, the scale could consist of items that differentiate the two groups.

Investigation of relationships between SVT and PVT scales revealed some notable associations. For parents in the clinical and control groups as negative appraisals of child or adolescent emotional and behavioral functioning on the BASC-2 increased, performance on PVTs dependent (in part) on working memory abilities decreased. Abnormally high BASC-2 F Index scores are designed to detect respondents that may attempt to make a child 'look bad' on a questionnaire. However, some items within that scale are associated with difficulties that children with impairments in working memory or attention may exhibit (i.e., *acts without thinking*, *forgets things*; Reynolds & Kamphaus, 2004). Thus, these associations are expected given overlap of items designed to ask about real-world experiences. Importantly, the parent BASC-2 F Index in the clinical and control group was not elevated to the degree of invalidity. In contrast, the parent BRIEF Negativity scale was not associated with child or adolescent PVT

performance. Similarly, SVTs from parents in the simulation group were not associated with child or adolescent PVTs. As indicated in the paragraph above, these results may suggest that investigation of an additional validity scale more specific to neuropsychological contexts of misrepresentation may be warranted in order to more accurately detect over-reporting of cognitive and somatic symptoms typically seen those contexts rather than more general contexts.

Additionally, there is a broad literature suggesting a low level of parent and child/adolescent agreement or association between the parent and self-report questionnaires (e.g., see Reynolds & Kamphaus, 2010). Correlational analyses were conducted for the clinical scales that overlapped in parent- and self-report questionnaires (see Appendix V). Observed associations were generally consistent with correlations published in the BASC-2 manual supplement (Reynolds & Kamphaus, 2010). Despite discrepant reports of emotional functioning, it may be valuable to consider and discuss the unique contribution or perspective of these two types of report in validity research. For example, when evaluating a younger child, it may be appropriate to place more weight on parent report and critically evaluate validity and clinical elevations due to developmental considerations associated with younger children and their emerging ability to fully report, recall, and describe experiences. Conversely, when evaluating an older adolescent, it may be more valuable to critically consider their self-report given emerging autonomy and ability to engage in research and clinical contexts. In summary, psychometric properties of the BASC-2 will be important to consider in future validity studies.

**Limitations and Future Directions**

The present study was limited by several factors and there remains opportunity for future development in multiple areas. Notably, each PVT was not incorporated into batteries for the medical-center participants. Thus, further exploration of the AST with adjusted cut-off scores and WRAML-2 Verbal Learning Forced Choice task with children and adolescents with various neurological, medical, and developmental conditions is warranted. Given promising sensitivity and specificity of these two measures further investigation is needed to more comprehensively document clinical utility. Further, the AST cut-off scores proposed in the present study using a community-recruited clinical group should be evaluated in samples where processing speed may be slowed (e.g., depression) to further evaluate the risk of false positives. While the proposed PVT and SVT cut-off scores have been systematically derived in the present study, they should be considered in other clinical contexts and populations, particularly within disability evaluations. The present study provides preliminary cut-off score recommendations for youth VSVT, RDS-R, and WRAML-2 embedded measures and suggests adjustments to youth DS ACSS, RDS, Rey FIT, and AST cut-offs. However, additional validation would increase confidence in utility of these PVTs across samples.

The present study also highlighted an important finding, that some PVTs (e.g., VSVT difficult items, embedded Digit Span PVTs) are strongly associated with working memory abilities. Additionally, working memory is known to be affected by multiple neurological (e.g., TBI, epilepsy) and developmental conditions (e.g., ADHD). Related to the current finding, adult VSVT studies reported that patients with intractable epilepsy exhibited a high rate of false positives when utilizing VSVT difficult item cut-off scores

derived from adult mTBI or non-clinical simulation samples (Grote et al., 2000; Loring et al., 2005). Further, epilepsy patients with low intellectual functioning and poor working memory displayed a higher risk for VSVT difficult item failure at certain cut-offs (Keary et al., 2013). These examples further highlight the need for consideration of diverse clinical samples in PVT cut-off score validation, so that clinicians may make informed decisions regarding their application. The present study, which included a mixed clinical sample, consisted of a small number of individuals with intellectual disability and/or epilepsy thus, it would be valuable to explore the VSVT in those specific populations with a larger sample size.

To further investigate the relationship between working memory and select PVT performances, exploratory analyses were conducted with a sub-sample of participants previously diagnosed with ADHD ($n$ = 60). When considering previously proposed cut-off scores from other samples (i.e., mTBI), children with ADHD exhibited a similar frequency of failure as the general clinical sample (see Appendix W), which was greater than the generally accepted false positive error rate of 10%, on measures such as the VSVT, embedded Digit Span measures, and AST tasks. When considering the newly proposed cut-off scores presented in the current study (see Table 10), participants with ADHD exhibited failure rates (see Appendix X) within a generally expected range. However, children with ADHD had unacceptable failure rates on the VSVT tasks (see Appendix X). While specificity was optimized in the larger clinical sample, this smaller subset of clinical patients, children and adolescents with ADHD, exhibited a higher rate of failure than ideal, thus further exploration of the VSVT and consideration of alternative cut-off scores is warranted (as suggested above). These exploratory analyses

suggest that while many of the proposed PVT cut-off scores are appropriate to apply when evaluating a child or adolescent with ADHD, the proposed VSVT cut-off scores should be conservatively considered. Further, while previous literature does not suggest that the VSVT performance is associated with sustained attention (see Brooks, 2012), there may be other cognitive factors such as impulsivity that negatively affect VSVT performance that warrant investigation.

The present study also identified that a PVT that employs a response speed component (AST) demonstrated the some of the most optimal classification statistics in the present study. Memory-like paradigms [e.g., CARB, MSVT, MVP, Rey FIT, TOMM (Tombaugh, 1996); VSVT, and WMT], developed initially for adults, are popular and have varying degrees of empirical support with children and adolescents (Kirkwood, 2015). However, it is possible that simple timed tasks may be more beneficial for detection of invalid performance in youth samples and warrant further exploration, given the promising results from the AST in the present study. Further, as presented above, when cut-off scores are identified for a broader clinical sample, or specifically children and adolescents with ADHD, for the AST there is increased confidence in limitation of false positive rates for speeded tasks.

In addition to future exploration of simple timed tasks, PVT researchers may wish to consider investigation of adaptive platforms or flexible approaches for youth. For example, adult researchers have identified that an abbreviated TOMM administration (e.g., first 10 items) can exhibit better sensitivity than traditional TOMM administration (Denning, 2014), which could lead to improved efficiency and accuracy of validity assessment. Researchers may wish to explore similar options in youth. If an abbreviated

administration (e.g., first block of the VSVT) indicates valid performance, then perhaps validity testing may be discontinued. Conversely, if invalidity is indicated, additional measures may be warranted to increase confidence in validity determination.

Consistent with the concept of increased confidence of validity determination, adult PVT and SVT literature highlights the importance of defining validity on the basis of multiple PVT and/or SVT failure. Slick and colleagues (1999) identified that two or more validity tests must be failed as a part of the criteria of Malingered Neurocognitive Dysfunction. Additionally within adult literature, methods for objectively identifying or interpreting failure of multiple validity indicators has emerged (see Odland, Lammy, Martin, Grote,& Mittenberg, 2015). Additional exploratory analyses were conducted using the community-recruited sample to evaluate multiple PVT failure rate for 14 possible PVTs (see Appendix Y). Of note, for tasks that included multiple opportunities for failure (i.e., embedded Digit Span measures) only one possible failure was considered. Simulators were much more likely to exhibit multiple PVT failure (> 2), than control or clinical participants. However, there were still clinical participants ($n$ = 5) that failed more than 2 PVTs. Though, all PVTs were included in that analysis and as the present study suggests, some PVTs may be more optimal than others for children and adolescents. In consideration of that point, statistical measures that take into account the sensitivity and specificity of each PVT or SVT failure in aggregation in order to improve detection of invalidity (versus false positive over-identification) have been proposed in adult literature (see Larrabee, 2008; Meyers et al., 2014). Thus, statistical evaluation of aggregated PVT failure is the logical next step in research for child and adolescent PVT

and SVT research in order to increase confidence in invalidity detection through utilization of measures with optimal classification statistics.

An additional limitation of the present study was that the researchers were not blind to group assignment. A survey of pediatric neuropsychologists indicated that some clinicians believe PVTs are unnecessary because they can determine subjectively that children and adolescents displayed invalid performance (Brooks et al., 2016). Due to the necessary study design, we were unable to effectively test if researchers could discriminate simulators from control or clinical participants. This remains an important area of further exploration. If experienced clinicians can effectively distinguish valid versus invalid presentations through interaction then use of SVTs and PVTs may not be critical; however, it would be highly valuable to test these assumptions in research and clinical contexts. Developmental literature does suggest that children and adolescents are capable of conscious deception that evades detection. Further, unconscious invalid presentations can create unclear profiles. Regardless of subjective appraisal of validity, some assert that objective PVTs and SVTs can be critical for clinical decision making (e.g., pre-surgical evaluations; Connery & Suchy, 2015). Further, discussing PVT or SVT failure with a parent, and, at times, the child/adolescent could result in meaningful conversations regarding psychosocial stressors as opposed to inaccurately attributing neurocognitive symptoms to an erroneous diagnosis. The ultimate goal is to utilize PVTs and SVTs as tools to inform clinical decision-making and improve patient outcomes.

In summary, this study identified that youth are capable of simulating cognitive difficulties and multiple PVTs could detect invalid performance. Parents are also able to exaggerate symptom report without detection from currently available SVTs. Application

of previously identified adult or youth mTBI PVT cut-off scores results in high false positive rates for most measures, particularly in younger children and those with intellectual disability. Thus, PVT cut-off scores should be adjusted for mixed clinical samples in order to prevent over identification of invalidity while maintaining moderate sensitivity. This study also provides a framework for future investigation and development of youth PVTs and SVTs that may exhibit strong clinical utility, important considerations for specific clinical samples, and suggests opportunities for aggregated evaluation of validity failure. Finally, responsible utilization of PVTs and SVTs could improve clinical decision-making and outcomes by providing additional clarity of validity and platforms for discussion of invalid profiles.

# REFERENCES


Abdelmoula, M., Chakroun, W., & Akrout, F. (2015). The effect of sample size and the number of items on reliability coefficients: Alpha and rho: A meta-analysis. *International Journal of Numerical Methods and Applications, 13,* 1-20. doi:10.17654/IJNMAMar2015_001_020

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB '97: Manual for the computerized assessment of response bias.* Durham, NC: Cogni-Syst.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders fifth edition (DSM-5).* Arlington, VA: American Psychiatric Association Publishing.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82.

Araujo, G. C., Antonini, T. N., Monahan, K., Gelfius, C., Klamar, K., Potts, M., Yeates, K. O., & Bodin, D. (2014). The relationship between suboptimal effort and post-concussion symptoms in children and adolescents with mild traumatic brain injury. *The Clinical Neuropsychologist, 28,*785-801. doi:10.1080/13854046.2014.896415

Axelrod, B., Fitchenberg, N., Millis, S., & Wertheimer, J. (2006). Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale – Third Edition. *The Clinical Neuropsychologist, 20,* 513-523.

Baer, R., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308-320. doi: 10.1207/S15327752JP720213

Baker, D. A., Connery, A. K., Kirk, J. W., & Kirkwood, M. W. (2014). Embedded performance validity indicators within the California Verbal Learning Test, Children's Version. *The Clinical Neuropsychologist, 28,* 116-127. doi:10.1080/13854046.2013.858184

Ben-Porath, Y. S. & Tellegan, A. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.

Berry, D. T. R. & Nelson, N. W. (2010). DSM-5 and malingering: A modest proposal. *Psychological Injury and Law, 3,* 295-303. doi:10.1007/s12207-010-9087-7

Berry, D. T. R. & Schipper, L. J. (2008) Assessment of feigned cognitive impairment using standard neuropsychological tests. In R. Rogers (Eds.), *Clinical assessment of malingering and deception* (pp. 237-254). New York, NY: The Guilford Press.

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15,* 19-45. doi:10.1076/clin.15.1.19.1907

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological testing. *Journal of the International Neuropsychological Society, 18,* 632-642. doi:10.1017/S1355617712000252

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

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335-340.

Boone, K. B., Salazar, X., Lu, P.,Warner-Chacon, K., & Razani, J. (2002). The Rey 15-Item recognition trial: A technique to enhance sensitivity of the Rey 15-Item memorization test. *Journal of Clinical and Experimental Neuropsychology, 24,* 561-573. doi:1380/3395/022405-561

Bracken, B. A. (2007). *Bracken School Readiness Assessment – Third Edition: Examiner's Manual.* San Antonio, TX: NSC Pearson, Inc.

Brennan, A. M., & Gouvier, W. D. (2006). Are we honestly studying malingering? A profile and comparison of simulated and suspected malingerers. *Applied Neuropsychology, 13,* 1–11.

Brooks, B. L. (2012). Victoria Symptom Validity Test performance in children and adolescents with neurological disorders. *Archives of Clinical Neuropsychology, 27,* 858-868. doi:10.1093/arclin/acs087

Brooks, B. L. & Ploetz, D. M. (2015). Embedded performance validity on the CVLT-C for youth with neurological disorders. *Archives of Clinical Neuropsychology, 30,* 200-206. doi:10.1093/arclin/acv017

Brooks, B. L., Ploetz, D. M., & Kirkwood, M. W. (2016). A survey of neuropsychologists' use of validity tests with children and adolescents, *Child Neuropsychology, 22,* 1001-1020. doi:10.1080/09297049.2015.1075491

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice

issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20,* 419-426. doi:10.1016/j.acn.2005.02.002

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. and Kaemmer, B. (1989). *Manual for administration and scoring of the MMPI-2,* Minneapolis, MN: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A) manual*. Minneapolis, MN: University of Minnesota Press.

Carone, D. A. (2014). Young child with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: Implications for mild TBI. *The Clinical Neuropsychologist, 28,* 146-162. doi: 10.1080/13854046.2013.861019

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist,* 529-546. doi:10.1080/13854040701346104

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security disability psychological consultative examination. *Applied Neuropsychology,* 143-149. doi:10.1016/j.acn.2006.10.003

Chafetz, M. D. & Dufrene, M. (2014). Malingering-by-proxy: Need for child protection and guidance of reporting. *Child Abuse & Neglect, 38,* 1755-1765. doi:10.1016/j.chiabu.2014.08.015

Chafetz, M. & Prentkowski, E. (2011). A case of malingering by proxy in a social security disability psychological consultative examination. *Applied Neuropsychology, 18,* 143-149. doi:10.1080/09084282.2011.570619

Cohen, M. J. (1997). *Children's Memory Scale (CMS): Manual.* San Antonio, TX: The Psychological Corporation.

Connery, A. K. & Suchy, Y. (2015). Managing noncredible performance in pediatric clinical assessment. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort.* (145-163). New York, NY: The Guilford Press.

Constantinou, M. & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*(2), 81-90.

Conti, R. P. (2004). Malingered ADHD in adolescents diagnosed with conduct disorder: A brief note. *Psychological Reports, 94,* 987-988. doi:10.2466/pr0.94.3.987-988

Courtney, J. C., Dinkins, J. P., Allen, L. M., & Kuroski, K. (2003). Age related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*(2), 109-116.

Curtis, K. L., Greve, K. W., Bianchini, K. J., & Brennan, A. (2006). California Verbal Learning Test indicators of malingered neurocognitive dysfunction. *Assessment, 13*(1), 46-61. doi: doi:10.1177/1073191105285210

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *California Verbal Learning Test-Children's Version (CVLT-C).* San Antonio, TX: The Psychological Corporation

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *The California Verbal Learning Test* (2nd ed.; CVLT-II). San Antonio, TX: The Psychological Corporation.

Denning J. H. (2014). The efficiency and accuracy of the Test of Memory Malingering trial 1, errors on the first 10 items of the test of memory malingering, and five embedded measures in predicting invalid test performance. Archives of Clinical Neuropsychology, 29, 417-432. doi:10.1093/arclin/acs044

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

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11,* 221-227. doi:10.1080/09297040490917298

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

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

Flaro, L. & Boone, K. (2009). Using objective effort measures to detect noncredible cognitive test performance in children and adolescents. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of Malingering Casebook* (pp. 369-376). New York, NY: Psychology Press.

Flaro, L., Green, P., & Blaskewitz, N. (2007). The importance of symptom validity testing in children: WMT and MSVT. *Praxis der Rechtspsychologie, 17,* 125-139.

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology, 23,* 501-509. doi:10.1016/j.acn.2008.04.001

Gervais, R., Ben-Porath, Y., Wygant, D., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14,* 196-208.

Gervais, R., Wygant, D., Sellbom, M., & Ben-Porath, Y. (2011). Association between symptom validity test failure and scores on the MMPI-2-RF validity and substantive scales. *Journal of Personality Assessment, 93*, 508-517.

Gidley-Larson, J. C., Flaro, L., Peterson, R. L., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2015). The Medical Symptom Validity Test Measures effort not ability in children: A comparison between mild TBI and Fetal Alcohol Spectrum Disorder samples. *Archives of Clinical Neuropsychology, 30,* 192-199. doi:10.1093/arclin/acv012

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function (BRIEF) Professional Manual.* Lutz, FL: PAR, Inc.

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014) The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of Clinical and Experimental* Neuropsychology, 36, 261-267. doi:10.1080/13803395.2013.879096

Green, P. (2003). *Green's Word Memory Test (WMT) for Windows: User manual.* Edmonton, Canada: Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Windows: User Manual.* Edmonton, Canada: Green's Publishing.

Green, P. & Flaro, L. (2014). Results from three performance validity tests in children with intellectual disability. *Applied Neuropsychology: Child, 5*, 25-34. doi: 10.1080/21622965.2014.935378

Green, P., Flaro. L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, & NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In C.R. Reynolds & A. Horton (Eds.), *Detection of malingering during head injury litigation* (2nd ed., pp. 201-219). New York: Plenum Press.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224.

Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology, 22*(6), 709-719.

Gunn, D., Batchelor, J., & Jones, M. (2010). Detection of simulated memory impairment in 6- to 11-year-old children. *Child Neuropsychology, 16,* 105-118. doi:10.1080/09297040903352564

Harrison, A. G. & Edwards, M. J. (2010). Symptom exaggeration in post-secondary students: Preliminary base rates in a Canadian sample. *Applied Neuropsychology, 2,* 135-143. doi: 10.1080/09084281003715642

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2008). Identifying students feigning dyslexia: Preliminary findings and strategies for detection. *Dyslexia, 14,* 228-246. doi: 10.1002/dys.366

Harrison, A. G., Green, P., & Flaro, L. (2012). The importance of symptom validity testing in adolescents and young adults undergoing assessments for learning or attention difficulties. *Canadian Journal of School Psychology, 27,* 98-113. doi: 10.1177/0829573512437024

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. doi:10.1080/13854040903155063

Hoelzle, J. B., Nelson, N. W., & Arbisi, P. A. (2012). MMPI-2 and MMPI-2-Restructured Form validity scales: Complementary approaches to evaluate response validity. *Psychology Injury and Law, 5,* 174-191. doi: 10.1007/s12207-012-9139-2

Horner, M. D., VanKirk, K. K., Dismuke, C. E., Turner, T. H., & Muzzy, W. (2014). Inadequate effort on neuropsychological evaluation is associated with increased healthcare utilization. *The Clinical Neuropsychologist, 28,* 703-713. doi:10.1080/13854046.2014.925143

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression (2nd ed.).* New York, NY: Wiley Interscience.

Jones, A. (2013). Victoria Symptom Validity Test: Cutoff scores for psychometrically defined malingering groups in a military sample. *The Clinical Neuropsychologist, 27,* 1373-1394. doi:10.1080/13854046.2013.851740

Keary, T. A., Frazier, T.W., Belzile, C. J., Chapin, J. S., Naugle, R. I., & Najm, I. M. (2013).Working memory and intelligence are associated with Victoria Symptom Validity Test hard item performance in patients with intractable epilepsy. *Journal of the International Neuropsychological Society, 19,* 314–323.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M.W. (2011). Performance on the Test of Memory Malingering (TOMM) among

a large clinic-referred pediatric sample. *Child Neuropsychology, 3,* 242-254, doi:10.1080/09297049.2010.533166

Kirk, J. W., Hutaff-Lee, C. F., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild Traumatic Brain Injury. *Assessment, 5,* 562-569. doi:10.1177/1073191114520626

Kirkwood, M. W. (2012). Overview of tests and techniques to detect negative response bias in children. In E.M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 136-161). New York, NY: Oxford University Press, Inc.

Kirkwood, M. W. (2015). Review of pediatric performance and symptom validity tests. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort.* (79-106). New York, NY: The Guilford Press.

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2014a ). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology, 20,* 245-252. doi:10.1080/09297049.2012.759553

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26,* 377-384. doi:10.1093/arclin/acr040

Kirkwood, M. W. & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 5,* 860-872. doi: 10.1080/13854040903527287

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16,* 604-618. doi:10.1080/09297049.2010.495059

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014b). Postconcussive symptom exaggeration after pediatric mild Traumatic Brain Injury. *Pediatrics, 4,* 643-650. doi:10.1542/peds.2013-3195

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24,* 36-45. doi:10.1037/a0024628

Lange, R. T., Pancholi, S., Bhagwat, A., Anderson-Barnes, V., & French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U.S. military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 34,* 453–466. doi:10.1080/13803395.2011.648175

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625-631. doi:10.1017/S1355617712000240

Larrabee, G. J. (2008).Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666-679. doi:10.1080/13854040701494987

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.) New York, NY: Oxford University Press.

Libow, J. A. (2000). Child and adolescent illness falsification. *Pediatrics, 105*(2), 336-342.

Lichtenstein, J. D., Erdodi, L. A., & Linnea, K. S. (2017). Introducing a forced-choice recognition task to the California Verbal Learning Test – Children's Version. *Child Neuropsychology, 23*, 284-299. doi: 10.1080/09297049.2015.1135422

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2007). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* 145-151. doi:10.1111/j.1466-8238.2007.00358

Loring, D. W., Lee, G. P., & Meador, K. J. (2005). 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

Loughan, A. R., Perna, R., & Hertza, J. (2012). The value of the Wechsler intelligence Scale for Children-Fourth Edition Digit Span as an embedded measure of effort: An investigation into children with dual diagnoses. *Archives of Clinical Neuropsychology, 27*, 716-724.

Lu, P. H. & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-yr-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16*, 90-96. doi:10.1076/clin.16.1.90.8328

Lucio, E., Durán, C., Graham, J. R., & Ben-Porath, Y. (2002). Identifying faking bad on the Minnesota Multiphasic Personality Inventory – Adolescent with Mexican adolescents. *Assessment, 9*(1), 62-69.

MacAllister, W. S. & Vasserman, M. (2015). Ethical considerations in pediatric validity testing. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluation exaggeration, feigning, and noncredible effort.* New York, NY: Guilford Press.

Macmillan, N. A. & Creelman, C. D. (1991). *Detection theory: A user's guide* (2nd ed.) Cambridge, England: Cambridge University Press.

McCaffrey, R. J. & Lynch, J. K. (2017). The importance of performance validity assessment in pediatric neuropsychological evaluations. *Journal of Pediatric Neuropsychology, 3,* 106-113. doi: 10.1007/s40817-017-0031-6.

McKinzey, R. K., Prieler, J., & Raven, J. (2003). Detection of children's malingering on Raven's Standard Progressive Matrices. *British Journal of Clinical Psychology, 42,* 95-99.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30(5),* 525-564.

Meyers, J. E., Miller, R. M., Thompson, L. M., Scalese, A. M., Allred, B. C… Lee, A. J. (2014). Using likelihood ratios to detect invalid performance with performance validity measures. *Archives of Clinical Neuropsychology, 29,* 224-235. doi:10.1093/arclin/acu001

Morey, L (2007). *Personality Assessment Inventory (PAI) Techincal Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Nagle, A. M., Everhart, D. E., Durham, T. W., McCammon, S. L., & Walker, M. (2006). Deception strategies in children: Examination of forced choice recognition and verbal learning and memory techniques. *Archives of Clinical Neuropsychology, 21,* 777-785. doi: 10.1016/j.acn.2006.06.011

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

Perna, R. B. & Loughan, A. R. (2013). The influence of effort on neuropsychological performance in children: Is performance on the TOMM indicative of neuropsychological ability? *Applied Neuropsychology: Child, 3,* 31-37. doi:10.1080/21622965.2012.686339

Perna, R. B., Loughan, A. R., Hertza, J., & Segraves, K. (2014). The value of embedded measures in detecting suboptimal effort in children: An investigation in to the WISC-IV Digit Span and CMS Verbal Memory Subtest. *Applied Neuropsychology: Child., 3,* 45-51. doi:10.1080/21622965.2012.291067

Ploetz, D. M., Mazur-Mosiewicz, A., Kirkwood, M. W., Sherman, E. M., & Brooks, B. L. (2014). Performance on the Test of Memory Malingering in children with neurological conditions. *Child Neuropsychology,* doi:10.1080/09297049.2014.986446

Rambo, P. L., Callahan, J. L., Hogan, L. R., Hullman, S., & Wrape, E. (2015). Effort testing in children: Can cognitive and symptom validity measures differentiate malingered performances? *Applied Neuropsychology: Child, 4,* 1-8. doi:10.1080/21622965.2013.780238

Reese, C. S., Suhr, J. A., & Riddle, T. L. (2012). Exploration of malingering indices in the Wechsler Adult Intelligence Scale–Fourth Edition Digit Span subtest. Archives of Clinical Neuropsychology, 27, 176–181. doi:10.1093/arclin/acr117

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

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children – Second edition (BASC-2).* Circle Pines, MN: American Guidance Service.

Rogers, R. (2008). Researching response styles. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception 3rd Edition* (411-434). New York, NY: The Guilford Press.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13,* 255-274.

Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment, 67*(2), 244-257.

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms (SIRS) and professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Salekin, R. T., Kubak, F. A., & Lee, Z. (2008). Deception in children and adolescents. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 343-364). New York, NY: The Guilford Press.

Schoenberg, M. R., & Scott, J. G. (2011). The neuropsychology referral and answering the referral question. *The little black book of neuropsychology: A syndrome-based approach*. New York, NY: Springer.

Schwartz, E., Erlodi, L., Rodriguez, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A., & Roth, R. M. (2016). CVLT-II Forced Choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society, 22,* 851-858. doi:10.1017/S1355617716000746

Sheslow, D. & Adams, W. (2003). *Wide Range Assessment of Memory and Learning Second Edition administration and technical manual*. Lutz, FL: Psychological Assessment Resources.

Silk-Eglit, G. M., Lynch, J. K., & McCaffrey, R. J. (2016). Validation of Victoria Symptom Validity Test cutoff scores among mild Traumatic Brain Injury litigants using a known-groups design. *Archives of Clinical Neuropsychology, 31*(3), 231-245. doi: 10.1093/arclin/acv108

Slick, D. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. Sherman, & O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test (VSVT): Professional manual.* Lutz, FL: Psychological Assessment Resources, Inc. (PAR, Inc.).

Slick, D., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingering neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545-561. doi:10.1080/13854040802054151

Spencer, R., Axelrod, B. N., Drag, L. L., Waldron-Perrine, B., Pangilinan, P. H., & Bieliauskas, L. A. (2013). WAIS-IV Reliable Digit Span is no more accurate than age corrected scaled score as an indicator of invalid performance in a veteran sample undergoing evaluation for mTBI. *The Clinical Neuropsychologist, 27*(8), 1362-1372. doi: 10.1080/13854046.2013.845248

Spencer, R., Tree, H., Drag, L., Pangilinan, P., & Bieliauskas, L. (2010). *Extending Reliable Digit Span with the WAIS-IV sequencing task: Preliminary results.* Paper presented at the at the 8th annual meeting for the American Academy of Clinical Neuropsychology Conference, Chicago, IL.

Stein, L. A. R., Graham, J. R., & Williams, C. L. (1995). Detecting fake-bad MMPI-A profiles. *Journal of Personality Assessment, 65(3),* 415-427.

Strauss, E., Sherman, E. M, & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

Stutts, J. T., Hickey, S. E., & Kasdan, M. L. (2003). Malingering by Proxy: A form of pediatric condition falsification. *Developmental and Behavioral Pediatrics, 24*(4), 276-278.

Suhr, J., Hammers, D., Dobbins-Buckland, 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,* 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,* 189-207. doi: 10.1080/09084280701509083

Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics, (6th Ed.).* Boston, MA: Pearson Education, Inc.

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology, 43,* 804-810. doi:10.1037/0012-1649.43.3.804

Talwar, V. & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development, 26,* 436-444. doi:10.1080/01650250143000373

Tarescavage, A. M., Wygant, D. B., Roger O. Gervais, R. O. & Ben-Porath, Y. S. (2012) Association between the MMPI-2 Restructured Form (MMPI-2-RF) and Malingered Neurocognitive Dysfunction among non-head injury disability claimants. *The Clinical Neuropsychologist, 27,* 313-335, doi: 10.1080/13854046.2012.744099

Tombaugh, T. N. (1996). *The Test of Memory Malingering (TOMM).* Toronto, ON: Multi-Health Systems, Inc.

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

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology, 17,* 755-774. doi:10.1002/acp.914

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V): Technical Manual.* San Antonio, TX: NSC Pearson, Inc.

Welsh, A. J., Bender, A., Whitman, L. A., Vasserman, M., & MacAllister, W. S. (2012). Clinical utility of reliable digit span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27,* 735-741. doi:10.1093/arclin/acs063

Whitney, K., Davis, J., Shepard, P., & Herman, S. (2008). Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23,* 777-786.

Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). *The Clinical Neuropsychologist, 24*(1), 153-168. doi: 10.1080/13854040903107791

Young, J. C., Sawyer, R. J., Roper, B. L., & Baughman, B. C. (2012). Expansion and re-examination of digit span effort indices on the WAIS-IV. *The Clinical Neuropsychologist, 26,* 147-159. doi:10.1080/1385404.2011.647083

Yurdugul, H. (2008). Minimum sample size for cronbach's coefficient alpha: A monte-carlo study. *Hacettepe University Journal of Education, 35,* 397-405.

Appendix A
Child and adolescent performance and symptom validity test research studies

| Author(s) | Test(s) | Population | N | Age range | Cut Score(s) | Results | | | | Research Method |
|---|---|---|---|---|---|---|---|---|---|---|
| Araujo et al. (2014) | CMS (RDS) | mTBI | 382 | 8 – 16 | RDS ≤ 6 and/or ACSS ≤ 5 | 20% of sample classified as invalid performance and had higher rates of reported post-concussion symptoms and poorer performance on Trails A and B | | | | Known Groups Comparison |
| Baker, Connery, Kirk, & Kirkwood (2014) | CVLT-C MSVT | mTBI | 411 | 8 - 16 | | Sn | Sp | | | Known Groups Comparison |
| | | | | | CVLT-C RD = -0.5 | .55 | .91 | | | |
| | | | | | CVLT-C RD = -1.0 | .41 | .97 | | | |
| Blaskewitz, Merten, & Kathmann (2008) | MSVT TOMM FIT RDS | Non-Clinical Community Sample | 38 | 6 - 11 | FIT ≤ 7<br>MSVT ≤ 90%<br>TOMM < 45<br>RDS ≤ 6 | None of the controls failed the MSVT, TOMM, & Rey FIT<br>59% of the controls failed RDS<br>70 to 90% of simulators failed the MSVT, TOMM, & RDS<br>10% of simulators failed the Rey FIT | | | | Simulation Study |
| Brooks (2012) | VSVT | Mixed Clinical | 100 | 6 – 19 | Total Items < 30 | 95% exhibited a valid performance on Total Items | | | | Descriptive Study |
| | | | | | Easy Items < 16 | 97% exhibited a valid performance on Easy Items | | | | |
| | | | | | Hard Items < 16 | 84% exhibited a valid performance on Hard Items | | | | |
| Brooks & Ploetz (2015) | CVLT-C TOMM | Mixed Clinical | 294 | 5 - 16 | | Sn | Sp | PPP | NPP | Known Groups Comparison |
| | | | | | CVLT-C RD *z* ≤ -0.5 | .81 | .67 | | | |
| | | | | | CVLT-C RD *z* ≤ -3.0 | .44 | .90 | .16 | .97 | |
| Constantinou & McCaffrey (2003) | TOMM FIT | Cross Cultural Nonclinical | 128 | 5 - 12 | TOMM Trial 2 < 45 | 98% of children achieved a valid TOMM score | | | | Descriptive Study |
| | | | | | FIT cut-score not reported | Age and education effects present in children until age 10 | | | | |

Appendix A (continued)
Child and adolescent performance and symptom validity test research studies

| Author(s) | Test(s) | Population | N | Age range | Cut Score(s) | Results | | | | Research Method |
|---|---|---|---|---|---|---|---|---|---|---|
| Green, Kirk, Connery, Baker, & Kirkwood (2014) | FIT | mTBI | 319 | 8 - 17 | | Sn | Sp | | | Known Groups Comparison |
| | | | | | FIT Recall < 9 | .12 | .98 | | | |
| | | | | | Recognition < 26 | .55 | .91 | | | |
| Kirk et al. (2014) | BASC-2 Self-Report F Index | mTBI | 274 | 8 - 17 | __ | Invalid performance group established with MSVT failure. No relationship between invalid MSVT and BASC-2 | | | | Known Groups |
| Kirkwood et al. (2014) | Automatized Sequences | mTBI | 452 | 8 - 17 | | Sn | Sp | AUC | | Known Groups Comparison |
| | | | | | Alphabet ≥ 8 seconds | .50 | .91 | .73 | | |
| | | | | | Counting ≥ 6 seconds | .50 | .92 | .75 | | |
| | | | | | Days of Week ≥ 4 seconds | .31 | .96 | .77 | | |
| | | | | | Months of Year ≥ 10 seconds | .36 | .90 | .76 | | |
| | | | | | Total Time ≥ 27 seconds | .55 | .90 | .80 | | |
| Kirkwood et al. (2011) | RDS | mTBI | 274 | 8 - 16 | | Sn | | Sp | | Known Groups Comparison |
| | | | | | ACSS ≤ 5 | .51 | | .96 | | |
| | | | | | RDS ≤ 6 | .51 | | .92 | | |
| Loughan et al. (2012) | RDS | Mixed Clinical | 51 | -- | | Sn | | Sp | | Known Groups Comparison |
| | | | | | RDS ≤ 4 | .43 | | .91 | | |
| Lichtenstein et al. (2017) | CVLT-C (FCR-C) | Mixed Clinical | 72 | 6 - 15 | | Sn | | Sp | | Known Groups Comparison |
| | | | | | ≤ 14/15 | .31 | | .87 | | |
| | | | | | ≤ 13 | .15 | | .94 | | |
| Lu & Boone (2002) | FIT WRT Dot Counting b Test | mTBI | 1 | 9 | FIT < 8 WRT < 6 Other cut-offs not reported | Exhibited invalid performance on all four PVTs | | | | Descriptive Case Study |
| Perna et al. (2014) | RDS | Mixed Clinical | 75 | 6 - 18 | RDS ≤ 4 | Sn | | Sp | | Known Groups Comparison |
| | | | | | | .44 | | .94 | | |
| Welsh et al. (2012) | RDS | Epilepsy | 54 | 6 – 17 | | Sn | Sp | PPP | NPP | Known Groups Comparison |
| | | | | | RDS ≤ 6 | 1.00 | .71 | .26 | 1.00 | |

Appendix A (continued)
Child and adolescent performance and symptom validity test research studies

*Note:* Cut scores are reported as a value at or below the specified number indicates invalidity. Results were documented as reported in manuscripts. If sensitivity, specificity, positive predictive power, or negative predictive power were available these were reported. AUC = Area under the curve from receiver operating characteristic analysis; BASC-2 = Behavior Assessment System for Children – Second Edition Self Report; CMS = Children's Memory Scale; CVLT-C = California Verbal Learning Test – Children's Version; FIT = Fifteen Item Test; MSVT = Medical Symptom Validity Test; mTBI = mild Traumatic Brain Injury; RDS = Reliable Digit Span; TOMM = Test of Memory Malingering; VSVT = Victoria Symptom Validity Test; Sn = sensitivity; Sp = specificity; PPP = positive predictive power; NPP = negative predictive power

# Appendix B

## Child or Adolescent Simulation Instructions

"You did well on those last 3 tasks, and I can tell that you tried your best. Now, I will ask you to pretend that you recently had an accident playing sports or on the playground and that you hit your head. After you hit your head you weren't feeling good and got to stay home from school for a couple days.

When you get back to school, your friend tells you that if kids have trouble paying attention, remembering things, can't think as quickly as they used to, have headaches, or feel more irritable than normal, then they can get out of doing some assignments and may only have to go to classes for half of the day. You decide that this would really help you in school, and it could be fun to get out of some classes you don't like.

On these next texts, I want you to pretend that you have trouble paying attention, can't remember some things, and can't complete things as fast as you normally would.

So let's practice, I am going to read a short list of words and I want you to pretend that you have trouble paying attention or remember things. Ready?

Chicken, Horse, Pig, Duck

Now tell me as many words as you remember."

If the child completes the task slowly or "remembers" less than 4 words or incorrect words say, "Good job pretending!"

If the child completes the task quickly and recalls all 4 words say, "Remember I want you to pretend that you have trouble paying attention or remembering things." Then try practice again.

Ask child to tell you in their own words what the instructions are and record response.

Clarify instructions as necessary.

# Appendix C

## Parent/guardian simulation instructions

"You will be asked to complete two questionnaires. We want you to pretend that your child recently had a head injury playing sports or on the playground and has returned to school a couple days after their injury. You are concerned that your child might experience some difficulty with headaches or dizziness, paying attention, staying organized, remembering things, completing tasks quickly, and keeping his/her emotions under control when he/she returns to school. You believe that your child should get academic accommodations for assignments and testing in case he/she needs them. Please answer these questionnaires in a manner that would help your child get accommodations at school."

Ask parent to tell you in their own words what they should do. Clarify as needed.

# Appendix D

## Debriefing instructions for simulation group

"Thank you for participating in our study. The purpose of this study was to find out if *certain* tests can help us determine if children are not doing their best with paying attention or remembering, or if parents are not giving accurate reports. While we gave you a scenario in which someone might experience some trouble with thinking after a minor injury that may not actually occur in real life. We are working on developing tests that help us diagnose cognitive problems more accurately so that we can ensure children and teens receive the right kind of therapy or treatments for their difficulties. While we asked you to pretend in this research study, it is important that you always do your best on tests and provide honest answers on questionnaires. Do you have any questions?"

# Appendix E

## Manipulation check for simulation instructions

Briefly describe the directions you were given.
What were you were supposed to pretend in this study?

_____

_____

Please rate how successfully you were able to understand the directions for this study.

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Not At All | | | Somewhat | | Completely |

Please rate how successfully you were able to follow the directions for this study.

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | Not At All | | Somewhat | | |
| | | | Completely | | |

Appendix F

**Test Order: A**

- ☐ *Parents: BRIEF, BASC-2, Demographics Form*
- ☐ WISC-V Vocabulary
- ☐ WISC-V Matrix Reasoning
- ☐ Bracken Number Identification
- ☐ Victoria Symptom Validity Test
- ☐ WRAML-2 Verbal Learning Immediate Recall
- ☐ WISC-V Digit Span
- ☐ WRAML-2 Verbal Learning Delayed Recall
- ☐ Rey 15-Item Test
- ☐ Automatized Sequences
- ☐ WRAML-2 Verbal Learning Recognition
- ☐ BASC-2 Self Report
- ☐ WRAML-2 Verbal Learning Forced Choice

**Test Order: B**

- ☐ *Parents: BRIEF, BASC-2, Demographics Form*
- ☐ WISC-V Vocabulary
- ☐ WISC-V Matrix Reasoning
- ☐ Bracken Number Identification
- ☐ Victoria Symptom Validity Test
- ☐ WRAML-2 Verbal Learning Immediate Recall
- ☐ Rey 15-Item Test
- ☐ Automatized Sequences
- ☐ WRAML-2 Verbal Learning Delayed Recall
- ☐ WISC-V Digit Span
- ☐ WRAML-2 Verbal Learning Recognition
- ☐ BASC-2 Self Report
- ☐ WRAML-2 Verbal Learning Forced Choice

Appendix F continued

**Test Order: C**

- ☐ *Parents: BRIEF, BASC-2, Demographics Form*
- ☐ WISC-V Vocabulary
- ☐ WISC-V Matrix Reasoning
- ☐ Bracken Number Identification
- ☐ BASC-2 Self Report
- ☐ WRAML-2 Verbal Learning Immediate Recall
- ☐ WISC-V Digit Span
- ☐ WRAML-2 Verbal Learning Delayed Recall
- ☐ Rey 15-Item Test
- ☐ Automatized Sequences
- ☐ WRAML-2 Verbal Learning Recognition
- ☐ Victoria Symptom Validity Test
- ☐ WRAML-2 Verbal Learning Forced Choice

**Test Order: D**

- ☐ *Give to parents: BRIEF, BASC-2, Demographics Form*
- ☐ WISC-V Vocabulary
- ☐ WISC-V Matrix Reasoning
- ☐ Bracken Number Identification
- ☐ BASC-2 Self Report
- ☐ WRAML-2 Verbal Learning Immediate Recall
- ☐ Rey 15-Item Test
- ☐ Automatized Sequences
- ☐ WRAML-2 Verbal Learning Delayed Recall
- ☐ WISC-V Digit Span
- ☐ WRAML-2 Verbal Learning Recognition
- ☐ Victoria Symptom Validity Test
- ☐ WRAML-2 Verbal Learning Forced Choice

Appendix G
WRAML-2 Verbal Learning Forced Choice Recognition

*(Administer 10 minutes after recognition)*

**Remember that long list of words I had you learn? I want to see which words from the list you remember now. Was boat or cake on the list? Was _____ or _____ on the list?**

*(Continue with the rest of the list for the appropriate age below. You may prompt/encourage the examinee to take a guess if they are unsure of the answer. Instructions may be abbreviated when the examinee understands the task.)*

| **Was _____ or _____ on the list?** | | | Score | Distractor Type |
|---|---|---|---|---|
| **Boat** | or | Cake | 0 1 | C |
| **Hat** | or | Stove | 0 1 | C |
| Quiet | or | **Door** | 0 1 | A |
| **Flag** | or | Lamp | 0 1 | C |
| Dream | or | **Wood** | 0 1 | A |
| Apple | or | **Sand** | 0 1 | C |
| **Nail** | or | Brave | 0 1 | A |
| Cow | or | **Ear** | 0 1 | C |
| **Game** | or | Soft | 0 1 | A |
| **Ice** | or | Love | 0 1 | A |
| Ball | or | **Map** | 0 1 | C |
| **Comb** | or | Luck | 0 1 | A |
| Banana | or | **Tree** | 0 1 | C |
| ***Ages 9 to adult continue with words below*** | | | | |
| Sleep | or | **Lake** | 0 1 | A |
| **Page** | or | Rabbit | 0 1 | C |
| Loud | or | **Ant** | 0 1 | A |
| **Forced Choice Recognition Raw Score** | | | | |
| **Forced Choice Recognition Percent Correct (≤8 years = ____ / 13) (9+ years = ____ / 16)** | | | | |

Appendix H
BASC-2 Alpha Coefficient Reliabilities for Clinical Scales
In Community-Recruited Clinical and Control Groups

| | Child Self-Report ($n$ = 32) | Adolescent Self-Report ($n$ = 14) | Parent Report Child ($n$ = 32) | Parent Report Adolescent ($n$ = 14) |
|---|---|---|---|---|
| Composite Scales | | | | |
| School Problems | .64 | .57 | __ | __ |
| Inattention/Hyperactivity | .35 | .45 | __ | __ |
| Personal Adjustment | .68 | .71 | __ | __ |
| Internalizing Problems | .90 | .76 | .93 | .93 |
| Behavioral Symptoms Index | __ | __ | .91 | .88 |
| Adaptive Skills | __ | __ | .92 | .91 |
| Externalizing Problems | __ | __ | .92 | .94 |
| Clinical Scales | | | | |
| Attitude to School | .82 | .10 | __ | __ |
| Attitude to Teachers | .32 | .65 | __ | __ |
| Sensation Seeking | __ | .16 | __ | __ |
| Atypicality | .76 | .69 | .86 | .60 |
| Locus of Control | .42 | .07 | __ | __ |
| Social Stress | .80 | .75 | __ | __ |
| Anxiety | .78 | .56 | .90 | .92 |
| Depression | .61 | .78 | .85 | .72 |
| Sense of Inadequacy | .59 | .14 | __ | __ |
| Somatization | __ | .45 | .82 | .87 |
| Attention Problems | .07 | .18 | .01 | .50 |
| Hyperactivity | .37 | .76 | .92 | .87 |
| Aggression | __ | __ | .80 | .86 |
| Conduct Problems | __ | __ | .86 | .88 |
| Withdrawal | __ | __ | .58 | .20 |
| Adaptive Scales | | | | |
| Relations with Parents | .74 | .78 | __ | __ |
| Interpersonal Relations | .45 | .08 | __ | __ |
| Self-Esteem | .22 | .29 | __ | __ |
| Self-Reliance | .59 | .57 | __ | __ |
| Adaptability | __ | __ | .64 | .63 |
| Social Skills | __ | __ | .89 | .90 |
| Leadership | __ | __ | .87 | .86 |
| Activities of Daily Living | __ | __ | .15 | .39 |
| Functional Communication | __ | __ | .59 | .84 |
| Validity Scales | | | | |
| F Index | .48 | .19 | .59 | .73 |
| L Index | .46 | .06 | __ | __ |
| V Index | .05 | .00 | __ | __ |

*Note:* BASC-2 = Behavior Assessment System for Children – Second Edition.

# Appendix I

BRIEF Parent Report Cronbach's Alpha Reliability Coefficients for Control and Clinical Participants ($n$ = 46)

| Subscales | Cronbach's Alpha |
|---|---|
| Inhibit | .92 |
| Shift | .86 |
| Emotional Control | .86 |
| Initiate | .79 |
| Working Memory | .93 |
| Plan/Organize | .88 |
| Org. of Material | .78 |
| Monitor | .86 |
| Index Scales | |
| BRI | .93 |
| MI | .97 |
| GEC | .98 |
| Validity Scales | |
| Negativity | .82 |
| Inconsistency | .89 |

*Note:* BRIEF = Behavior Rating Index of Executive Function; BRI = Behavioral Regulation Index; MI = Metacognition Index; GEC = Global Executive Composite.

Appendix J
Control Group VSVT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| VSVT | | | | | | | | | |
| Easy Correct | ≤ 7 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 8 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.26% |
| | ≤ 10 | .11 | 1.00 | 0.00 | .89 | 1.00 | .56 | 0.00% | 10.53% |
| | ≤ 11 | .11 | 1.00 | 0.00 | .89 | 1.00 | .56 | 0.00% | 10.53% |
| | ≤ 12 | .16 | 1.00 | 0.00 | .84 | 1.00 | .58 | 0.00% | 15.79% |
| | ≤ 13 | .16 | 1.00 | 0.00 | .84 | 1.00 | .58 | 0.00% | 15.79% |
| | ≤ 14 | .21 | 1.00 | 0.00 | .79 | 1.00 | .59 | 0.00% | 21.05% |
| | ≤ 15 | .26 | 1.00 | 0.00 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 16 | .26 | 1.00 | 0.00 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 17 | .37 | 1.00 | 0.00 | .63 | 1.00 | .65 | 0.00% | 36.84% |
| | ≤ 18 | .37 | 1.00 | 0.00 | .63 | 1.00 | .65 | 0.00% | 36.84% |
| | ≤ 19 | .37 | 1.00 | 0.00 | .63 | 1.00 | .65 | 0.00% | 36.84% |
| | ≤ 20 | .42 | .95 | 9.26 | .61 | .89 | .66 | 4.55% | 42.11% |
| | **≤ 21** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | ≤ 22 | .47 | .86 | 3.47 | .61 | .75 | .66 | 18.18% | 47.37% |
| | ≤ 23 | .68 | .77 | 3.01 | .41 | .72 | .74 | 22.72% | 68.42% |
| VSVT | | | | | | | | | |
| Difficult Correct | ≤ 7 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 8 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.26% |
| | ≤ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.26% |
| | ≤ 10 | .12 | 1.00 | 0.00 | .89 | 1.00 | .56 | 0.00% | 10.53% |
| | ≤ 11 | .21 | 1.00 | 0.00 | .79 | 1.00 | .59 | 0.00% | 21.05% |
| | ≤ 12 | .32 | 1.00 | 0.00 | .68 | 1.00 | .63 | 0.00% | 31.58% |
| | ≤ 13 | .32 | 1.00 | 0.00 | .68 | 1.00 | .63 | 0.00% | 31.58% |
| | ≤ 14 | .42 | .95 | 9.26 | .61 | .89 | .66 | 4.54% | 42.11% |
| | ≤ 15 | .47 | .95 | 10.42 | .55 | .90 | .68 | 4.54% | 47.37% |
| | ≤ 16 | .47 | .95 | 10.42 | .55 | .90 | .68 | 4.54% | 47.37% |
| | ≤ 17 | .47 | .95 | 10.42 | .55 | .90 | .68 | 4.54% | 47.37% |
| | **≤ 18** | **.53** | **.95** | **11.58** | **.50** | **.91** | **.70** | **4.54%** | **52.63%** |
| | ≤ 19 | .63 | .77 | 2.78 | .48 | .71 | .71 | 22.72% | 63.16% |
| | ≤ 20 | .68 | .68 | 2.15 | .46 | .65 | .71 | 31.81% | 68.42% |
| | ≤ 21 | .68 | .64 | 1.88 | .50 | .62 | .70 | 36.36% | 68.42% |
| | ≤ 22 | .79 | .50 | 1.58 | .42 | .58 | .73 | 50.00% | 78.95% |
| | ≤ 23 | .84 | .27 | 1.16 | .58 | .50 | .67 | 72.73% | 84.21% |

Appendix G continued on the next page

Appendix J (continued)
Control Group VSVT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| VSVT | | | | | | | | | |
| Total Correct | ≤ 22 | .05 | 1.00 | 0 | .95 | 1.00 | .55 | 0.00% | 5.26% |
| | ≤ 23 | .16 | 1.00 | 0 | .84 | 1.00 | .58 | 0.00% | 15.79% |
| | ≤ 24 | .16 | 1.00 | 0 | .84 | 1.00 | .58 | 0.00% | 15.79% |
| | ≤ 25 | .26 | 1.00 | 0 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 26 | .26 | 1.00 | 0 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 27 | .26 | 1.00 | 0 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 28 | .26 | 1.00 | 0 | .74 | 1.00 | .61 | 0.00% | 26.32% |
| | ≤ 29 | .32 | 1.00 | 0 | .68 | 1.00 | .63 | 0.00% | 31.58% |
| | ≤ 30 | .32 | 1.00 | 0 | .68 | 1.00 | .63 | 0.00% | 31.58% |
| | ≤ 31 | .42 | 1.00 | 0 | .58 | 1.00 | .67 | 0.00% | 42.11% |
| | ≤ 32 | .42 | 1.00 | 0 | .58 | 1.00 | .67 | 0.00% | 42.11% |
| | ≤ 33 | .47 | 1.00 | 0 | .53 | 1.00 | .69 | 0.00% | 47.37% |
| | **≤ 34** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 35** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 36** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 37** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 38** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 39** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | **≤ 40** | **.47** | **.95** | **10.42** | **.55** | **.90** | **.68** | **4.55%** | **47.37%** |
| | ≤ 41 | .53 | .86 | 3.86 | .55 | .77 | .68 | 13.64% | 52.63% |
| | ≤ 42 | .58 | .82 | 3.18 | .51 | .73 | .69 | 18.18% | 57.89% |
| | ≤ 43 | .68 | .73 | 2.51 | .43 | .68 | .73 | 27.27% | 68.42% |
| | ≤ 44 | .68 | .68 | 2.15 | .46 | .65 | .71 | 31.82% | 68.42% |
| | ≤ 45 | .74 | .64 | 2.03 | .41 | .64 | .74 | 36.36% | 73.68% |
| | ≤ 46 | .79 | .45 | 1.45 | .46 | .56 | .71 | 54.55% | 78.95% |
| | ≤ 47 | .84 | .27 | 1.16 | .58 | .50 | .67 | 72.73% | 84.21% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; VSVT = Victoria Symptom Validity Test

Appendix K
Combined Clinical Groups VSVT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| VSVT | | | | | | | | | |
| Easy Correct | ≤ 7 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 8 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 9 | .05 | .99 | 3.89 | .96 | .33 | .89 | 1.35% | 5.26% |
| | ≤ 10 | .11 | .99 | 7.79 | .91 | .50 | .90 | 1.35% | 10.53% |
| | ≤ 11 | .11 | .99 | 7.79 | .91 | .50 | .90 | 1.35% | 10.53% |
| | ≤ 12 | .16 | .97 | 5.84 | .87 | .43 | .90 | 2.70% | 15.79% |
| | ≤ 13 | .16 | .97 | 4.67 | .87 | .38 | .90 | 3.38% | 15.79% |
| | ≤ 14 | .21 | .97 | 6.23 | .82 | .44 | .91 | 3.38% | 21.05% |
| | ≤ 15 | .26 | .95 | 5.56 | .77 | .42 | .91 | 4.73% | 26.32% |
| | ≤ 16 | .26 | .95 | 4.87 | .78 | .38 | .91 | 5.41% | 26.32% |
| | **≤ 17** | **.37** | **.93** | **5.45** | **.68** | **.41** | **.92** | **6.76%** | **36.84%** |
| | ≤ 18 | .37 | .93 | 4.96 | .68 | .39 | .92 | 7.43% | 36.84% |
| | ≤ 19 | .37 | .90 | 3.64 | .70 | .32 | .92 | 10.14% | 36.84% |
| | ≤ 20 | .42 | .88 | 3.46 | .66 | .31 | .92 | 12.16% | 42.11% |
| | ≤ 21 | .47 | .83 | 2.80 | .63 | .26 | .92 | 16.89% | 47.37% |
| | ≤ 22 | .47 | .72 | 1.71 | .73 | .18 | .91 | 27.70% | 47.37% |
| | ≤ 23 | .68 | .55 | 1.53 | .57 | .16 | .93 | 44.59% | 68.42% |
| VSVT | | | | | | | | | |
| Difficult Correct | ≤ 7 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | .54 | 0.00% | 0.00% |
| | ≤ 8 | .05 | .99 | 3.89 | .96 | .33 | .89 | 1.35% | 5.26% |
| | ≤ 9 | .05 | .98 | 2.60 | .97 | .25 | .89 | 2.03% | 5.26% |
| | ≤ 10 | .11 | .97 | 3.89 | .92 | .33 | .89 | 2.70% | 10.53% |
| | ≤ 11 | .21 | .96 | 5.19 | .82 | .40 | .90 | 4.05% | 21.05% |
| | **≤ 12** | **.32** | **.91** | **3.60** | **.75** | **.32** | **.91** | **8.78%** | **31.58%** |
| | ≤ 13 | .32 | .91 | 3.33 | .76 | .30 | .91 | 9.46% | 31.58% |
| | ≤ 14 | .42 | .85 | 2.83 | .68 | .27 | .92 | 14.86% | 42.11% |
| | ≤ 15 | .47 | .81 | 2.50 | .65 | .24 | .92 | 18.92% | 47.37% |
| | ≤ 16 | .47 | .77 | 2.06 | .68 | .21 | .92 | 22.97% | 47.37% |
| | ≤ 17 | .47 | .76 | 1.95 | .70 | .20 | .92 | 24.32% | 47.37% |
| | ≤ 18 | .53 | .72 | 1.85 | .66 | .19 | .92 | 28.38% | 52.63% |
| | ≤ 19 | .63 | .66 | 1.83 | .56 | .19 | .93 | 34.46% | 63.16% |
| | ≤ 20 | .68 | .55 | 1.51 | .58 | .16 | .93 | 45.27% | 68.42% |
| | ≤ 21 | .68 | .48 | 1.32 | .66 | .14 | .92 | 52.03% | 68.42% |
| | ≤ 22 | .79 | .33 | 1.18 | .64 | .13 | .92 | 66.89% | 78.95% |
| | ≤ 23 | .84 | .18 | 1.02 | .90 | .12 | .90 | 82.43% | 84.21% |

Appendix H continued on next page

Appendix K (continued)
Combined Clinical Groups VSVT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| VSVT | | | | | | | | | |
| Total Correct | ≤ 21 | 0.00 | .99 | 0.00 | 1.01 | 0.00 | .89 | 0.68% | 0.00% |
| | ≤ 22 | .05 | .99 | 3.89 | .96 | .33 | .89 | 1.35% | 5.26% |
| | ≤ 23 | .16 | .98 | 7.79 | .86 | .50 | .90 | 2.03% | 15.79% |
| | ≤ 24 | .16 | .96 | 3.89 | .88 | .33 | .90 | 4.05% | 15.79% |
| | ≤ 25 | .26 | .96 | 6.49 | .77 | .45 | .91 | 4.05% | 26.32% |
| | ≤ 26 | .26 | .96 | 6.49 | .77 | .45 | .91 | 4.05% | 26.32% |
| | ≤ 27 | .26 | .95 | 4.87 | .78 | .38 | .91 | 5.41% | 26.32% |
| | ≤ 28 | .26 | .94 | 4.33 | .78 | .36 | .91 | 6.08% | 26.32% |
| | ≤ 29 | .32 | .94 | 5.19 | .73 | .40 | .91 | 6.08% | 31.58% |
| | ≤ 30 | .32 | .93 | 4.67 | .73 | .38 | .91 | 6.76% | 31.58% |
| | **≤ 31** | **.42** | **.90** | **4.15** | **.64** | **.35** | **.92** | **10.14%** | **42.11%** |
| | ≤ 32 | .42 | .89 | 3.89 | .65 | .33 | .92 | 10.81% | 42.11% |
| | ≤ 33 | .47 | .89 | 4.12 | .59 | .35 | .93 | 11.49% | 47.37% |
| | ≤ 34 | .47 | .88 | 3.89 | .60 | .33 | .93 | 12.16% | 47.37% |
| | ≤ 35 | .47 | .86 | 3.51 | .61 | .31 | .93 | 13.51% | 47.37% |
| | ≤ 36 | .47 | .85 | 3.19 | .62 | .29 | .93 | 14.86% | 47.37% |
| | ≤ 37 | .47 | .84 | 2.92 | .63 | .27 | .93 | 16.22% | 47.37% |
| | ≤ 38 | .47 | .79 | 2.26 | .67 | .22 | .92 | 20.95% | 47.37% |
| | ≤ 39 | .47 | .76 | 1.95 | .70 | .20 | .92 | 24.32% | 47.37% |
| | ≤ 40 | .47 | .73 | 1.75 | .72 | .18 | .92 | 27.03% | 47.37% |
| | ≤ 41 | .53 | .71 | 1.81 | .67 | .19 | .92 | 29.05% | 52.63% |
| | ≤ 42 | .58 | .64 | 1.61 | .66 | .17 | .92 | 35.82% | 57.89% |
| | ≤ 43 | .68 | .59 | 1.66 | .54 | .18 | .94 | 41.22% | 68.42% |
| | ≤ 44 | .68 | .49 | 1.35 | .64 | .15 | .92 | 50.68% | 68.42% |
| | ≤ 45 | .74 | .43 | 1.28 | .62 | .14 | .93 | 57.43% | 73.68% |
| | ≤ 46 | .79 | .30 | 1.12 | .71 | .13 | .92 | 70.27% | 78.95% |
| | ≤ 47 | .84 | .16 | 1.01 | .97 | .11 | .89 | 83.78% | 84.21% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; VSVT = Victoria Symptom Validity Test.

Appendix L
Control Group Embedded Digit Span Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Digit Span | | | | | | | | | |
| DS ACSS | ≤ 1 | .35 | 1.00 | 0.00 | .65 | 1.00 | .63 | 0.00% | 35.00% |
| | ≤ 2 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≤ 3 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | ≤ 4 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | ≤ 5 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | ≤ 6 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | **≤ 7** | **.55** | **.95** | **12.10** | **.47** | **.92** | **.70** | **4.55%** | **55.00%** |
| | ≤ 8 | .60 | .82 | 3.30 | .49 | .75 | .69 | 18.18% | 60.00% |
| | ≤ 9 | .75 | .77 | 3.30 | .32 | .75 | .77 | 22.72% | 75.00% |
| | ≤ 10 | .85 | .55 | 1.87 | .28 | .63 | .80 | 45.45% | 85.00% |
| | ≤ 11 | .90 | .45 | 1.65 | .22 | .60 | .83 | 54.55% | 90.00% |
| | ≤ 12 | .90 | .27 | 1.24 | .37 | .53 | .75 | 72.73% | 90.00% |
| | ≤ 13 | .95 | .09 | 1.05 | .55 | .49 | .67 | 90.91% | 95.00% |
| RDS | ≤ 2 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≤ 3 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≤ 4 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≤ 5 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | **≤ 6** | **.50** | **.96** | **11.50** | **.52** | **.91** | **.69** | **4.35%** | **50.00%** |
| | ≤ 7 | .65 | .78 | 2.99 | .45 | .72 | .72 | 21.74% | 65.00% |
| | ≤ 8 | .80 | .57 | 1.84 | .35 | .62 | .76 | 65.22% | 80.00% |
| | ≤ 9 | 1.00 | .35 | 1.53 | 0 | .57 | 1.00 | 86.96% | 100% |
| RDS-R | ≤ 4 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≤ 5 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≤ 6 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≤ 7 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≤ 8 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | ≤ 9 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | **≤ 10** | **.50** | **.91** | **5.50** | **.55** | **.83** | **.67** | **9.09%** | **50.00%** |
| | ≤ 11 | .60 | .86 | 4.40 | .46 | .80 | .70 | 13.64% | 60.00% |
| | ≤ 12 | .80 | .64 | 2.20 | .31 | .67 | .78 | 36.36% | 80.00% |
| | ≤ 13 | .85 | .36 | 1.34 | .41 | .55 | .73 | 63.64% | 85.00% |
| | ≤ 14 | .95 | .09 | 1.05 | .55 | .49 | .67 | 90.91% | 95.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; DS ACSS = Digit Span Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span Revised

Appendix M
Combined Clinical Groups Embedded Digit Span Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Digit Span | | | | | | | | | |
| DS ACSS | ≤ 1 | .35 | .99 | 50.75 | .65 | .88 | .92 | 0.69% | 35.00% |
| | ≤ 2 | .40 | .99 | 29.00 | .61 | .80 | .92 | 1.38% | 40.00% |
| | ≤ 3 | .45 | .97 | 13.05 | .57 | .64 | .93 | 3.45% | 45.00% |
| | **≤ 4** | **.50** | **.92** | **6.59** | **.54** | **.48** | **.93** | **7.59%** | **50.00%** |
| | ≤ 5 | .50 | .85 | 3.30 | .59 | .31 | .92 | 15.17% | 50.00% |
| | ≤ 6 | .50 | .74 | 1.91 | .68 | .21 | .91 | 26.20% | 50.00% |
| | ≤ 7 | .55 | .59 | 1.33 | .77 | .15 | .90 | 41.38% | 55.00% |
| | ≤ 8 | .60 | .48 | 1.14 | .84 | .14 | .90 | 52.41% | 60.00% |
| | ≤ 9 | .75 | .36 | 1.17 | .70 | .14 | .91 | 64.14% | 75.00% |
| | ≤ 10 | .85 | .23 | 1.10 | .66 | .13 | .92 | 77.24% | 85.00% |
| | ≤ 11 | .90 | .16 | 1.07 | .63 | .13 | .92 | 84.14% | 90.00% |
| | ≤ 12 | .90 | .09 | .99 | 1.12 | .12 | .87 | 91.03% | 90.00% |
| | ≤ 13 | .95 | .04 | .99 | 1.21 | .12 | .86 | 95.86% | 95.00% |
| | ≤ 14 | .95 | .03 | .98 | 1.81 | .12 | .80 | 97.24% | 95.00% |
| | ≤ 15 | 1.00 | .01 | 1.01 | 0 | .12 | 1.00 | 98.62% | 100% |
| | ≤ 16 | 1.00 | .01 | 1.01 | 0 | .12 | 1.00 | 99.31% | |
| RDS | ≤ 2 | .15 | .99 | 21.30 | .86 | .75 | .89 | 0.70% | 15.00% |
| | ≤ 3 | .15 | .99 | 21.30 | .86 | .75 | .89 | 0.70% | 15.00% |
| | ≤ 4 | .40 | .99 | 28.60 | .61 | .80 | .92 | 1.41% | 40.00% |
| | **≤ 5** | **.45** | **.92** | **5.81** | **.60** | **.45** | **.92** | **7.75%** | **45.00%** |
| | ≤ 6 | .50 | .80 | 2.54 | .62 | .26 | .92 | 19.72% | 50.00% |
| | ≤ 7 | .65 | .56 | 1.47 | .63 | .17 | .92 | 44.37% | 65.00% |
| | ≤ 8 | .84 | .32 | 1.23 | .50 | .14 | .94 | 68.31% | 80.00% |
| | ≤ 9 | 1.00 | .14 | 1.16 | 0.00 | .14 | 1.00 | 85.92% | 100% |
| | ≤ 10 | 1.00 | .05 | 1.05 | 0.00 | .13 | 1.00 | 95.07% | 100% |
| | ≤ 11 | 1.00 | .01 | 1.01 | 0.00 | .13 | 1.00 | 96.48% | 100% |
| | ≤ 12 | 1.00 | .02 | 1.02 | 0.00 | .13 | 1.00 | 97.89% | 100% |
| | ≤ 13 | 1.00 | .01 | 1.01 | 0.00 | .13 | 1.00 | 98.59% | 100% |
| RDS-R | ≤ 2 | 0.00 | .99 | 0.00 | 1.01 | 0.00 | .86 | 0.81% | 0.00% |
| | ≤ 3 | 0.00 | .99 | 0.00 | 1.01 | 0.00 | .86 | 0.81% | 0.00% |
| | ≤ 4 | .15 | .99 | 18.45 | .86 | .75 | .88 | 0.81% | 15% |
| | ≤ 5 | .20 | .98 | 8.20 | .82 | .57 | .88 | 2.44% | 20% |
| | ≤ 6 | .30 | .96 | 7.38 | .73 | .55 | .89 | 4.07% | 30% |
| | ≤ 7 | .40 | .95 | 8.20 | .63 | .57 | .91 | 4.88% | 40% |
| | **≤ 8** | **.50** | **.92** | **6.15** | **.54** | **.50** | **.92** | **8.13%** | **50%** |
| | ≤ 9 | .50 | .85 | 3.42 | .59 | .36 | .91 | 14.63% | 50% |
| | ≤ 10 | .50 | .72 | 1.76 | .70 | 22 | .90 | 28.46% | 50% |
| | ≤ 11 | .60 | .61 | 1.54 | .66 | .20 | .90 | 39.02% | 60% |
| | ≤ 12 | .80 | .42 | 1.39 | .47 | .18 | .93 | 57.72% | 80% |
| | ≤ 13 | .85 | .29 | 1.20 | .51 | .16 | .92 | 70.73% | 85% |
| | ≤ 14 | .95 | .15 | 1.12 | .32 | .15 | .95 | 84.55% | 95% |
| | ≤ 15 | .65 | .07 | 1.02 | .77 | .14 | .89 | 93.50% | 95% |
| | ≤ 16 | 1.00 | .03 | 1.03 | 0 | .14 | 1.00 | 96.75% | 100% |
| | ≤ 17 | 1.00 | .03 | 1.03 | 0 | .14 | 1.00 | 96.75% | 100% |
| | ≤ 18 | 1.00 | .02 | 1.03 | 0 | .14 | 1.00 | 97.56% | 100% |
| | ≤ 19 | 1.00 | .02 | 1.02 | 0 | .14 | 1.00 | 98.37% | 100% |
| | ≤ 20 | 1.00 | .01 | 1.01 | 0 | .14 | 1.00 | 99.19% | 100% |

Appendix N
Control Group WRAML-2 Embedded Measures Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| WRAML-2 Discriminability | ≤ 27.50 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| | ≤ 50.00 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≤ 55.00 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≤ 55.88 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≤ 57.50 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≤ 60.00 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≤ 65.00 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | ≤ 75.00 | .55 | 1.00 | 0.00 | .45 | 1.00 | .71 | 0.00% | 55.00% |
| | **≤ 87.50** | **.60** | **.91** | **6.60** | **.44** | **.86** | **.71** | 9.09% | 60.00% |
| | ≤ 90.00 | .65 | .86 | 4.77 | .41 | .81 | .73 | 13.64% | 65.00% |
| | ≤ 92.50 | .65 | .64 | 1.79 | .55 | .62 | .67 | 36.36% | 65.00% |
| | ≤ 94.12 | .70 | .64 | 1.93 | .47 | .64 | .70 | 36.36% | 70.00% |
| | ≤ 95.00 | .70 | .45 | 1.28 | .66 | .54 | .63 | 54.55% | 70.00% |
| | ≤ 97.06 | .70 | .32 | 1.03 | .94 | .48 | .54 | 68.18% | 70.00% |
| | ≤ 97.50 | .75 | .14 | .87 | 1.83 | .44 | .38 | 86.36% | 75.00% |
| d Prime | ≤ -1.00 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| | ≤ -0.5 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≤ 0.0 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≤ 0.5 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≤ 1.0 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | ≤ 1.5 | .55 | 1.00 | 0.00 | .45 | 1.00 | .71 | 0.00% | 55.00% |
| | ≤ 2.0 | .55 | 1.00 | 0.00 | .45 | 1.00 | .71 | 0.00% | 55.00% |
| | ≤ 2.5 | .60 | .95 | 13.20 | .42 | .92 | .72 | 4.5% | 60.00% |
| | **≤ 3.0** | **.60** | **.91** | **6.60** | **.44** | **.86** | **.71** | **4.5%** | **60.00%** |
| | ≤ 3.5 | .65 | .64 | 1.79 | .55 | .62 | .67 | 13.6% | 65.00% |
| | ≤ 4.0 | .70 | .27 | .96 | 1.10 | .47 | .50 | 72.7% | 70.00% |
| | ≤ 4.5 | .75 | .14 | .87 | 1.83 | .44 | .38 | 86.4% | 75.00% |
| | ≤ 5.0 | 1.00 | 0.00 | 1.00 | 0.00 | .48 | 0.00 | 100% | 100% |
| Forced Choice % Correct | ≤ 38 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≤ 44 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≤ 50 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≤ 56 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≤ 63 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | ≤ 81 | .50 | 1.00 | 0.00 | .50 | 1.00 | .69 | 0.00% | 50.00% |
| | **≤ 92** | **.50** | **.95** | **11.00** | **.52** | **.91** | **.68** | **4.55%** | **50.00%** |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; WRAML-2 = Wide Range Assessment of Memory and Learning

Appendix O
Clinical Group WRAML-2 Embedded Measures Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| WRAML-2 Discriminability | ≤ 27.50 | .05 | 1.00 | 0.00 | .95 | 1.00 | .81 | 1.50% | 5.00% |
| | ≤ 50.00 | .10 | .99 | 8.10 | .91 | .67 | .82 | 1.50% | 10.00% |
| | ≤ 55.00 | .20 | .99 | 16.20 | .81 | .80 | .83 | 1.50% | 20.00% |
| | ≤ 55.88 | .25 | .99 | 20.25 | .76 | .83 | .84 | 1.50% | 25.00% |
| | ≤ 57.50 | .30 | .99 | 24.30 | .71 | .86 | .85 | 1.50% | 30.00% |
| | ≤ 60.00 | .40 | .98 | 16.20 | .62 | .80 | .87 | 3.00% | 40.00% |
| | ≤ 65.00 | .45 | .96 | 12.15 | .57 | .75 | .88 | 4.50% | 45.00% |
| | **≤ 75.00** | **.55** | **.95** | **11.14** | **.47** | **.73** | **.90** | **6.0%** | **55.00%** |
| | ≤ 87.50 | .60 | .83 | 3.47 | .48 | .46 | .89 | 20.9% | 60.00% |
| | ≤ 90.00 | .60 | .78 | 2.70 | .51 | .40 | .89 | 26.9% | 60.00% |
| | ≤ 92.50 | .65 | .73 | 2.93 | .48 | .37 | .89 | 32.8% | 65.00% |
| | ≤ 94.12 | .65 | .69 | 2.11 | .51 | .34 | .89 | 37.3% | 65.00% |
| | ≤ 95.00 | .65 | .60 | 1.65 | .58 | .29 | .88 | 47.8% | 65.00% |
| | ≤ 97.06 | .70 | .57 | 1.62 | .53 | .29 | .88 | 52.2% | 70.00% |
| | ≤ 97.50 | .75 | .37 | 1.19 | .68 | .23 | .86 | 76.1% | 75.00% |
| d Prime | ≤ -1.00 | .50 | 1.00 | 0.00 | .95 | 1.00 | .78 | 0.00% | 5.00% |
| | ≤ -0.5 | .10 | 1.00 | 0.00 | .90 | 1.00 | .79 | 0.00% | 10.00% |
| | ≤ 0.0 | .15 | 1.00 | 0.00 | .85 | 1.00 | .80 | 0.00% | 15.00% |
| | ≤ 0.5 | .30 | .99 | 20.10 | .71 | .86 | .83 | 1.50% | 30.00% |
| | ≤ 1.0 | .45 | .96 | 10.05 | .58 | .75 | .85 | 4.50% | 45.00% |
| | **≤ 1.5** | **.55** | **.93** | **7.37** | **.49** | **.69** | **.87** | **7.50%** | **55.00%** |
| | ≤ 2.0 | .55 | .90 | 5.26 | .50 | .61 | .87 | 10.40% | 55.00% |
| | ≤ 2.5 | .60 | .84 | 3.65 | .48 | .52 | .88 | 16.40% | 60.00% |
| | ≤ 3.0 | .60 | .75 | 2.36 | .54 | .41 | .86 | 25.40% | 60.00% |
| | ≤ 3.5 | .65 | .60 | 1.61 | .59 | .33 | .85 | 40.30% | 65.00% |
| | ≤ 4.0 | .70 | .40 | 1.17 | .74 | .26 | .82 | 59.70% | 70.00% |
| | ≤ 4.5 | .75 | .24 | .99 | 1.05 | .23 | .76 | 76.10% | 75.00% |
| | ≤ 5.0 | 1.00 | 0.00 | 1.00 | 0.00 | .23 | 0.00 | 100% | 100% |
| Forced Choice % | ≤ 38 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≤ 44 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≤ 50 | .25 | 1.00 | 0.00 | .75 | 1.00 | .61 | 0.00% | 25.00% |
| | ≤ 56 | .40 | 1.00 | 0.00 | .60 | 1.00 | .66 | 0.00% | 40.00% |
| | ≤ 63 | .45 | 1.00 | 0.00 | .55 | 1.00 | .68 | 0.00% | 45.00% |
| | ≤ 81 | .50 | 1.00 | 0.00 | .50 | 1.00 | .70 | 0.00% | 50.00% |
| | **≤ 92** | **.50** | **1.00** | **0.00** | **.50** | **1.00** | **.70** | **0.00%** | **50.00%** |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; WRAML-2 = Wide Range Assessment of Memory and Learning

Appendix P
Control Group Rey FIT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Rey FIT | | | | | | | | | |
| Recognition | ≤ 2 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| | ≤ 3 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≤ 4 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≤ 5 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | **≤ 6** | **.35** | **1.00** | **0.00** | **.65** | **1.00** | **.63** | **0.00%** | **35.00%** |
| | **≤ 7** | **.35** | **1.00** | **0.00** | **.65** | **1.00** | **.63** | **0.00%** | **35.00%** |
| | **≤ 8** | **.35** | **1.00** | **0.00** | **.65** | **1.00** | **.63** | **0.00%** | **35.00%** |
| | ≤ 9 | .35 | .86 | 2.57 | .75 | .70 | .59 | 13.64% | 35.00% |
| | ≤ 10 | .45 | .77 | 1.98 | .71 | .64 | .61 | 22.73% | 45.00% |
| | ≤ 11 | .50 | .77 | 2.20 | .65 | .67 | .63 | 22.73% | 50.00% |
| | ≤ 12 | .55 | .55 | 1.21 | .83 | .53 | .57 | 45.45% | 55.00% |
| | ≤ 13 | .65 | .45 | 1.19 | .77 | .52 | .59 | 54.55% | 65.00% |
| | ≤ 14 | .75 | .41 | 1.27 | .61 | .54 | .64 | 59.09% | 75.00% |
| False Positives | ≥ 1 | .45 | .85 | 3.00 | .65 | .75 | .61 | 13.64% | 45.00% |
| | **≥ 2** | **.30** | **.91** | **3.30** | **.77** | **.75** | **.59** | **13.64%** | **30.00%** |
| | ≥ 3 | .20 | .91 | 2.20 | .88 | .67 | .56 | 9.09% | 30.00% |
| | ≥ 4 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 5 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 20.00% |
| | ≥ 6 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≥ 7 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 15.00% |
| | ≥ 8 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≥ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; Rey FIT = Rey 15-Item Test

Appendix Q
Community Clinical Group Rey FIT Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Clinical | Simulation |
| Rey FIT | | | | | | | | | |
| Recognition | ≤ 2 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| | **≤ 3** | **.20** | **.91** | **2.30** | **.88** | **.67** | **.57** | **8.70%** | **20.00%** |
| | ≤ 4 | .20 | .87 | 1.53 | .92 | .57 | .56 | 13.00% | 20.00% |
| | ≤ 5 | .25 | .87 | 1.92 | .86 | .63 | .57 | 13.00% | 25.00% |
| | ≤ 6 | .35 | .87 | 2.68 | .75 | .70 | .61 | 13.00% | 35.00% |
| | ≤ 7 | .35 | .87 | 2.68 | .75 | .70 | .61 | 13.00% | 35.00% |
| | ≤ 8 | .35 | .83 | 2.01 | .79 | .64 | .59 | 17.40% | 35.00% |
| | ≤ 9 | .35 | .78 | 1.61 | .83 | .58 | .58 | 21.70% | 35.00% |
| | ≤ 10 | .45 | .74 | 1.73 | .74 | .60 | .61 | 26.10% | 45.00% |
| | ≤ 11 | .50 | .74 | 1.92 | .68 | .63 | .63 | 26.10% | 50.00% |
| | ≤ 12 | .55 | .70 | 1.81 | .65 | .61 | .64 | 30.40% | 55.00% |
| | ≤ 13 | .65 | .70 | 2.14 | .50 | .65 | .70 | 30.40% | 65.00% |
| | ≤ 14 | .25 | .57 | .58 | 1.33 | .33 | .46 | 56.50% | 75.00% |
| False Positives | ≥ 1 | .45 | .74 | 1.73 | .74 | .60 | .61 | 26.09% | 45.00% |
| | **≥ 2** | **.30** | **.96** | **6.90** | **.73** | **.86** | **.61** | **4.35%** | **30.00%** |
| | ≥ 3 | .20 | .96 | 4.60 | .84 | .80 | .58 | 4.35% | 20.00% |
| | ≥ 4 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≥ 5 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 15.00% |
| | ≥ 6 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 15.00% |
| | ≥ 7 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 10.00% |
| | ≥ 8 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; Rey FIT = Rey 15-Item Test

Appendix R
Control Group Automatized Sequences Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Control | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Automatized Sequences | | | | | | | | | |
| Alphabet Time | ≥ 3 | 1.00 | .05 | 1.05 | 0.00 | .49 | 1.00 | 95.45% | |
| | ≥ 4 | 1.00 | .14 | 1.16 | 0.00 | .51 | 1.00 | 86.36% | 100% |
| | ≥ 5 | .85 | .32 | 1.25 | .47 | .53 | .70 | 68.18% | 85.00% |
| | ≥ 6 | .80 | .45 | 1.47 | .44 | .57 | .71 | 54.55% | 80.00% |
| | ≥ 7 | .70 | .73 | 2.57 | .41 | .70 | .73 | 27.27% | 80.00% |
| | ≥ 8 | .70 | .82 | 3.85 | .37 | .78 | .75 | 18.18% | 70.00% |
| | ≥ 9 | .60 | .86 | 4.40 | .46 | .80 | .70 | 18.18% | 70.00% |
| | ≥ 10 | .60 | .86 | 4.40 | .46 | .80 | .70 | 18.18% | 70.00% |
| | ≥ 11 | .60 | .86 | 4.40 | .46 | .80 | .70 | 18.18% | 70.00% |
| | ≥ 12 | .60 | .86 | 4.40 | .46 | .80 | .70 | 18.18% | 70.00% |
| | ≥ 13 | .60 | .86 | 4.40 | .46 | .80 | .70 | 13.64% | 60.00% |
| | **≥ 14** | **.55** | **.91** | **6.05** | **.50** | **.85** | **.69** | **9.09%** | **55.00%** |
| | ≥ 15 | .50 | .95 | 11.00 | .52 | .91 | .68 | 4.55% | 50.00% |
| | ≥ 16 | .45 | 1.00 | 0.00 | .55 | 1.00 | .68 | 0.00% | 45.00% |
| | ≥ 17 | .45 | 1.00 | 0.00 | .55 | 1.00 | .68 | 0.00% | 45.00% |
| | ≥ 18 | .45 | 1.00 | 0.00 | .55 | 1.00 | .68 | 0.00% | 45.00% |
| | ≥ 19 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≥ 20 | .35 | 1.00 | 0.00 | .65 | 1.00 | .63 | 0.00% | 35.00% |
| | ≥ 21 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≥ 22 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 27 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 31 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≥ 42 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| Counting Time | ≥ 3 | 1.00 | .05 | 1.05 | 0.00 | .49 | 1.00 | 95.45% | 100% |
| | ≥ 4 | 1.00 | .09 | 1.10 | 0.00 | .50 | 1.00 | 90.91% | 100% |
| | ≥ 5 | .85 | .45 | 1.56 | .33 | .59 | .77 | 54.55% | 85.00% |
| | ≥ 6 | .80 | .77 | 3.52 | .26 | .76 | .81 | 22.73% | 80.00% |
| | **≥ 7** | **.65** | **.95** | **14.3** | **.37** | **.93** | **.75** | **4.55%** | **65.00%** |
| | ≥ 8 | .55 | 1.00 | 0.00 | .45 | 1.00 | .71 | 0.00% | 55.00% |
| | ≥ 11 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0.00% | 45.00% |
| | ≥ 12 | .40 | 1.00 | 0.00 | .60 | 1.00 | .65 | 0.00% | 40.00% |
| | ≥ 13 | .35 | 1.00 | 0.00 | .65 | 1.00 | .63 | 0.00% | 35.00% |
| | ≥ 16 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≥ 18 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 23 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≥ 25 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≥ 54 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| Days of the Week | ≥ 2 | .95 | .05 | 1.00 | 1.10 | .48 | .50 | 95.45% | 95.00% |
| | ≥ 3 | .85 | .64 | 2.34 | .24 | .68 | .82 | 36.36% | 85.00% |
| | ≥ 4 | .65 | .86 | 4.77 | .41 | .81 | .73 | 13.64% | 65.00% |
| | **≥ 5** | **.50** | **1.00** | **0.00** | **.50** | **1.00** | **.69** | **0%** | **50.00%** |
| | ≥ 6 | .45 | 1.00 | 0.00 | .55 | 1.00 | .67 | 0% | 45.00% |
| | ≥ 7 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0% | 30.00% |

Appendix R continued on the next page

Appendix R (continued)
Control Group Automatized Sequences Classification Statistics

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Days of the Week | ≥ 8 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 9 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 10 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 11 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≥ 17 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |
| Months of Year | ≥ 4 | 1.00 | .09 | 1.10 | 0.00 | .50 | 1.00 | 90.91% | 100% |
| | ≥ 5 | .90 | .59 | 2.20 | .17 | .67 | .87 | 40.91% | 90.00% |
| | ≥ 6 | .75 | .68 | 2.36 | .37 | .68 | .75 | 31.82% | 75.00% |
| | ≥ 7 | .65 | .73 | 2.39 | .48 | .68 | .70 | 27.27% | 75.00% |
| | ≥ 8 | .60 | .73 | 2.20 | .55 | .67 | .67 | 27.27% | 65.00% |
| | ≥ 9 | .60 | .77 | 2.64 | .52 | .71 | .68 | 22.73% | 65.00% |
| | ≥ 10 | .60 | .82 | 3.30 | .49 | .75 | .69 | 22.73% | 60.00% |
| | ≥ 11 | .40 | .82 | 2.20 | .73 | .67 | .60 | 18.18% | 60.00% |
| | ≥ 12 | .40 | .86 | 2.93 | .69 | .73 | .61 | 18.18% | 40.00% |
| | ≥ 16 | .35 | .86 | 2.57 | .75 | .70 | .59 | 18.18% | 35.00% |
| | ≥ 17 | .30 | .86 | 2.2 | .81 | .67 | .58 | 18.18% | 30.00% |
| | ≥ 18 | .20 | .86 | 1.47 | .93 | .57 | .54 | 18.18% | 20.00% |
| | ≥ 22 | .15 | .86 | 1.10 | .98 | .50 | .53 | 13.64% | 15.00% |
| | **≥ 23** | **.10** | **.91** | **1.1** | **.99** | **.50** | **.53** | **9.09%** | **10.00%** |
| | ≥ 26 | .05 | .91 | .55 | 1.05 | .33 | .51 | 9.09% | 10.00% |
| | ≥ 28 | .05 | .95 | 1.10 | 1.00 | .50 | .53 | 4.55% | 5.00% |
| | ≥ 35 | 0.00 | .95 | 0.00 | 1.05 | 0.00 | .51 | 4.55% | 0.00% |
| Total Time | ≥ 12 | 1.00 | .05 | 1.05 | 0.00 | .49 | 1.00 | 95.45% | 100% |
| | ≥ 13 | 1.00 | .09 | 1.10 | 0.00 | .50 | 1.00 | 90.91% | 100% |
| | ≥ 14 | 1.00 | .23 | 1.29 | 0.00 | .54 | 1.00 | 77.27% | 100% |
| | ≥ 15 | .90 | .27 | 1.24 | .37 | .53 | .75 | 72.73% | 100% |
| | ≥ 16 | .90 | .32 | 1.32 | .31 | .55 | .78 | 68.18% | 100% |
| | ≥ 17 | .90 | .41 | 1.52 | .24 | .58 | .82 | 59.09% | 100% |
| | ≥ 18 | .90 | .50 | 1.80 | .02 | .62 | .85 | 50.00% | 90.00% |
| | ≥ 20 | .85 | .64 | 2.34 | .24 | .68 | .82 | 36.36% | 85.00% |
| | ≥ 21 | .85 | .68 | 2.67 | .22 | .71 | .83 | 31.82% | 85.00% |
| | ≥ 22 | .75 | .68 | 2.36 | .37 | .68 | .75 | 31.82% | 75.00% |
| | ≥ 24 | .75 | .77 | 3.30 | .32 | .75 | .77 | 22.73% | 75.00% |
| | ≥ 25 | .65 | .77 | 2.86 | .45 | .72 | .71 | 22.73% | 70.00% |
| | ≥ 33 | .60 | .77 | 2.64 | .52 | .71 | .68 | 22.73% | 60.00% |
| | ≥ 34 | .60 | .82 | 3.30 | .49 | .75 | .69 | 18.18% | 60.00% |
| | ≥ 36 | .55 | .82 | 3.03 | .55 | .73 | .67 | 18.18% | 55.00% |
| | ≥ 38 | .50 | .82 | 2.75 | .61 | .71 | .64 | 18.18% | 50.00% |
| | ≥ 39 | .50 | .86 | 3.67 | .58 | .77 | .66 | 13.64% | 50.00% |
| | **≥ 45** | **.50** | **.91** | **5.50** | **.55** | **.83** | **.67** | **9.09%** | **50.00%** |
| | ≥ 47 | .45 | .91 | 4.95 | .61 | .82 | .65 | 9.09% | 45.00% |
| | ≥ 52 | .40 | .91 | 4.40 | .66 | .80 | .63 | 9.09% | 40.00% |
| | ≥ 58 | .35 | .95 | 7.70 | .68 | .88 | .62 | 4.55% | 35.00% |
| | ≥ 59 | .30 | 1.00 | 0.00 | .70 | 1.00 | .61 | 0.00% | 30.00% |
| | ≥ 62 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 82 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 86 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≥ 92 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| | ≥ 123 | .05 | 1.00 | 0.00 | .95 | 1.00 | .54 | 0.00% | 5.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power

Appendix S
Community Clinical Group Automatized Sequences Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Automatized Sequences | | | | | | | | | |
| Alphabet Time | ≥ 2 | 1.00 | 0.00 | 1.00 | 0.00 | .48 | 0.00 | 100% | |
| | ≥ 3 | 1.00 | .04 | 1.04 | 0.00 | .48 | 1.00 | 95.65% | |
| | ≥ 4 | 1.00 | .13 | 1.15 | 0.00 | .50 | 1.00 | 86.96% | 100% |
| | ≥ 5 | .85 | .30 | 1.22 | .49 | .52 | .70 | 65.22% | 85.00% |
| | ≥ 6 | .80 | .57 | 1.84 | .35 | .62 | .76 | 43.48% | 80.00% |
| | ≥ 7 | .70 | .65 | 2.01 | .46 | .64 | .71 | 34.78% | 70.00% |
| | ≥ 8 | .70 | .74 | 2.68 | .41 | .70 | .74 | 26.09% | 70.00% |
| | ≥ 9 | .60 | .74 | 2.30 | .54 | .67 | .68 | 26.09% | 60.00% |
| | ≥ 10 | .60 | .78 | 2.76 | .51 | .71 | .69 | 21.74% | 60.00% |
| | ≥ 11 | .60 | .83 | 3.45 | .48 | .75 | .70 | 17.39% | 60.00% |
| | ≥ 12 | .60 | .87 | 4.60 | .46 | .80 | .71 | 13.04% | 60.00% |
| | ≥ 13 | .60 | .87 | 4.60 | .46 | .80 | .71 | 13.04% | 60.00% |
| | ≥ 14 | .55 | .87 | 4.21 | .52 | .79 | .69 | 13.04% | 55.00% |
| | ≥ 15 | .50 | .87 | 3.83 | .58 | .77 | .67 | 13.04% | 50.00% |
| | ≥ 16 | .45 | .87 | 3.45 | .63 | .75 | .65 | 13.04% | 45.00% |
| | ≥ 18 | .40 | .87 | 3.07 | .69 | .73 | .63 | 13.04% | 40.00% |
| | **≥ 19** | **.40** | **.91** | **4.60** | **.66** | **.80** | **.64** | **8.70%** | **40.00%** |
| | ≥ 20 | .35 | .91 | 4.03 | .71 | .78 | .62 | 8.70% | 35.00% |
| | ≥ 21 | .30 | .91 | 3.45 | .77 | .75 | .60 | 8.70% | 30.00% |
| | ≥ 22 | .25 | .96 | 5.75 | .78 | .83 | .59 | 4.35% | 25.00% |
| | ≥ 27 | .20 | .96 | 4.60 | .84 | .80 | .58 | 4.35% | 20.00% |
| | ≥ 31 | .15 | .96 | 3.45 | .89 | .75 | .56 | 4.35% | 15.00% |
| | ≥ 40 | .05 | .96 | 1.15 | .99 | .50 | .54 | 4.35% | 5.00% |
| | ≥ 42 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |
| Counting Time | ≥ 4 | 1.00 | .09 | 1.10 | 0.00 | .49 | 1.00 | 91.30% | 100% |
| | ≥ 5 | .85 | .30 | 1.22 | .49 | .52 | .70 | 69.57% | 85.00% |
| | ≥ 6 | .80 | .65 | 2.30 | .31 | .67 | .79 | 34.78% | 80.00% |
| | ≥ 7 | .65 | .83 | 3.74 | .42 | .76 | .73 | 17.39% | 65.00% |
| | **≥ 8** | **.55** | **.91** | **6.33** | **.49** | **.85** | **.70** | **8.70%** | **55.00%** |
| | ≥ 11 | .45 | .91 | 5.18 | .60 | .82 | .66 | 8.70% | 45.00% |
| | ≥ 12 | .40 | .91 | 4.6 | .66 | .80 | .64 | 8.70% | 40.00% |
| | ≥ 13 | .35 | .91 | 4.03 | .71 | .78 | .62 | 8.70% | 35.00% |
| | ≥ 14 | .30 | .96 | 6.90 | .73 | .86 | .61 | 4.35% | 30.00% |
| | ≥ 16 | .30 | 1.00 | 0.00 | .70 | 1.00 | .62 | 0.00% | 30.00% |
| | ≥ 18 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≥ 23 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 15.00% |
| | ≥ 25 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 54 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |
| Days of the Week | ≥ 1 | .95 | 0.00 | .95 | 0.00 | .45 | 0.00 | 100% | 100% |
| | ≥ 2 | .95 | .04 | .99 | 1.15 | .46 | .50 | 95.65% | 95.00% |
| | ≥ 3 | .85 | .43 | 1.50 | .35 | .57 | .77 | 56.52% | 85.00% |
| | ≥ 4 | .65 | .78 | 2.99 | .45 | .72 | .72 | 21.74% | 65.00% |
| | ≥ 5 | .50 | .87 | 3.83 | .58 | .77 | .67 | 13.04% | 50.00% |
| | **≥ 6** | **.45** | **.96** | **10.35** | **.58** | **.90** | **.67** | **4.35%** | **45.00%** |
| | ≥ 7 | .30 | .96 | 6.90 | .73 | .86 | .61 | 4.35% | 30.00% |
| | ≥ 8 | .25 | .96 | 5.75 | .78 | .83 | .59 | 4.35% | 25.00% |

Appendix S (continued)
Community Clinical Group Automatized Sequences Classification Statistics

| PVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score: Clinical | Cumulative % of sample at cut-score: Simulation |
|---|---|---|---|---|---|---|---|---|---|
| Days of the Week | ≥ 9 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 10 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≥ 11 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 17 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |
| Months of Year | ≥ 4 | 1.00 | .04 | 1.05 | 0.00 | .48 | 1.00 | 95.65% | 100% |
| | ≥ 5 | .90 | .26 | 1.22 | .38 | .51 | .75 | 73.91% | 90.00% |
| | ≥ 6 | .75 | .30 | 1.08 | .82 | .48 | .58 | 69.57% | 75.00% |
| | ≥ 7 | .65 | .48 | 1.25 | .73 | .52 | .61 | 52.17% | 65.00% |
| | ≥ 8 | .65 | .48 | 1.25 | .73 | .52 | .61 | 52.17% | 65.00% |
| | ≥ 9 | .60 | .52 | 1.25 | .77 | .52 | .60 | 47.83% | 60.00% |
| | ≥ 10 | .60 | .52 | 1.25 | .77 | .52 | .60 | 47.83% | 60.00% |
| | ≥ 12 | .40 | .52 | .84 | 1.15 | .42 | .50 | 47.83% | 40.00% |
| | ≥ 16 | .30 | .78 | 1.38 | .89 | .55 | .56 | 21.74% | 35.00% |
| | ≥ 17 | .25 | .83 | 1.43 | .91 | .56 | .56 | 17.39% | 30.00% |
| | ≥ 18 | .20 | .87 | 1.53 | .92 | .57 | .56 | 13.04% | 20.00% |
| | ≥ 21 | .15 | .87 | 1.15 | .98 | .50 | .54 | 13.04% | 15.00% |
| | **≥ 22** | **.15** | **.91** | **1.73** | **.93** | **.60** | **.55** | **8.70%** | **15.00%** |
| | ≥ 23 | .10 | .91 | 1.15 | .99 | .50 | .54 | 8.70% | 10.00% |
| | ≥ 25 | .05 | .91 | .58 | 1.04 | .33 | .53 | 8.70% | 5.00% |
| | ≥ 28 | .05 | .96 | 1.15 | .99 | .50 | .54 | 4.35% | 5.00% |
| | ≥ 44 | .00 | .96 | 0.00 | 1.05 | 0.00 | .52 | 4.35% | 0.00% |
| Total Time | ≥ 14 | 1.00 | .09 | 1.10 | 0.00 | .49 | 1.00 | 91.30% | 100% |
| | ≥ 15 | .90 | .17 | 1.09 | .58 | .49 | .67 | 82.61% | 100% |
| | ≥ 16 | .90 | .26 | 1.22 | .38 | .51 | .75 | 73.91% | 100% |
| | ≥ 17 | .90 | .26 | 1.22 | .38 | .51 | .75 | 73.91% | 100% |
| | ≥ 18 | .90 | .35 | 1.38 | .29 | .55 | .80 | 65.22% | 90.00% |
| | ≥ 20 | .85 | .35 | 1.30 | .43 | .53 | .73 | 65.22% | 85.00% |
| | ≥ 22 | .75 | .39 | 1.23 | .64 | .52 | .64 | 60.87% | 75.00% |
| | ≥ 23 | .75 | .43 | 1.33 | .58 | .54 | .67 | 56.52% | 75.00% |
| | ≥ 25 | .70 | .48 | 1.34 | .63 | .54 | .65 | 52.17% | 70.00% |
| | ≥ 26 | .60 | .52 | 1.25 | .77 | .52 | .60 | 47.83% | 60.00% |
| | ≥ 28 | .60 | .61 | 1.53 | .66 | .57 | .64 | 39.13% | 60.00% |
| | ≥ 29 | .60 | .65 | 1.73 | .61 | .60 | .65 | 34.78% | 60.00% |
| | ≥ 33 | .60 | .78 | 2.76 | .51 | .71 | .69 | 21.74% | 60.00% |
| | ≥ 36 | .55 | .78 | 2.53 | .58 | .69 | .67 | 21.74% | 55.00% |
| | ≥ 37 | .55 | .78 | 2.53 | .58 | .69 | .67 | 21.74% | 55.00% |
| | ≥ 38 | .50 | .83 | 2.88 | .61 | .71 | .66 | 17.39% | 50.00% |
| | ≥ 47 | .50 | .87 | 3.83 | .58 | .77 | .67 | 13.04% | 45.00% |
| | ≥ 49 | .40 | .87 | 3.07 | .69 | .73 | .63 | 13.04% | 40.00% |
| | **≥ 52** | **.40** | **.91** | **4.60** | **.66** | **.80** | **.64** | **8.70%** | **40.00%** |
| | ≥ 58 | .35 | .91 | 4.03 | .71 | .78 | .62 | 8.70% | 35.00% |
| | ≥ 59 | .30 | .91 | 3.45 | .77 | .75 | .60 | 8.70% | 30.00% |
| | ≥ 62 | .25 | .91 | 2.88 | .82 | .71 | .58 | 8.70% | 25.00% |
| | ≥ 69 | .20 | .91 | 2.30 | .88 | .67 | .57 | 8.70% | 20.00% |
| | ≥ 82 | .20 | .96 | 4.60 | .84 | .80 | .58 | 4.35% | 20.00% |
| | ≥ 86 | .15 | .96 | 3.45 | .89 | .75 | .56 | 4.35% | 15.00% |
| | ≥ 92 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 123 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |

Appendix T
Control Group SVT Classification Statistics

| SVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Control | Simulation |
| BASC-2 Child | | | | | | | | | |
| V Index | ≥ 1 | .35 | .91 | 3.85 | .72 | .78 | .61 | 9.09% | 100.00% |
| | **≥ 2** | **.35** | **.95** | **7.70** | **.68** | **.88** | **.62** | **4.55%** | **35.00%** |
| | ≥ 3 | .25 | 1.00 | 0.00 | .75 | 1.00 | .59 | 0.00% | 25.00% |
| | ≥ 4 | .20 | 1.00 | 0.00 | .80 | 1.00 | .58 | 0.00% | 20.00% |
| | ≥ 5 | .15 | 1.00 | 0.00 | .85 | 1.00 | .56 | 0.00% | 15.00% |
| | ≥ 9 | .10 | 1.00 | 0.00 | .90 | 1.00 | .55 | 0.00% | 10.00% |
| BASC-2 Parent | | | | | | | | | |
| F Index | ≥ 1 | .70 | .87 | 5.37 | .35 | .82 | .77 | 13.04% | 70.00% |
| | **≥ 3** | **.40** | **1.00** | **0.00** | **.60** | **1.00** | **.66** | **0.00%** | **40.00%** |
| | ≥ 4 | .35 | 1.00 | 0.00 | .65 | 1.00 | .64 | 0.00% | 35.00% |
| | ≥ 5 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≥ 9 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 12 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |
| BRIEF - Parent | | | | | | | | | |
| Negativity | **≥ 1** | **.75** | **.91** | **8.63** | **.27** | **.88** | **.81** | **8.70%** | **75.00%** |
| | ≥ 2 | .60 | 1.00 | 0.00 | .40 | 1.00 | .74 | 0.00% | 60.00% |
| | ≥ 3 | .55 | 1.00 | 0.00 | .45 | 1.00 | .72 | 0.00% | 55.00% |
| | ≥ 4 | .40 | 1.00 | 0.00 | .60 | 1.00 | .66 | 0.00% | 40.00% |
| | ≥ 5 | .35 | 1.00 | 0.00 | .65 | 1.00 | .64 | 0.00% | 35.00% |
| | ≥ 6 | .30 | 1.00 | 0.00 | .70 | 1.00 | .62 | 0.00% | 30.00% |
| | ≥ 7 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 15.00% |
| | ≥ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function

Appendix U
Community Clinical Group SVT Classification Statistics

| SVT | Cut-Score | Sn | Sp | LR+ | LR- | PPP | NPP | Cumulative % of sample at cut-score | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Clinical | Simulation |
| BASC-2 Child | | | | | | | | | |
| V Index | ≥ 1 | .35 | .87 | 2.68 | .75 | .70 | .61 | 13.04% | 100.00% |
| | **≥ 2** | **.35** | **.91** | **4.03** | **.71** | **.78** | **.62** | **8.70%** | **35.00%** |
| | ≥ 3 | .25 | 1.00 | 0.00 | .75 | 1.00 | .61 | 0.00% | 25.00% |
| | ≥ 4 | .20 | 1.00 | 0.00 | .80 | 1.00 | .59 | 0.00% | 20.00% |
| | ≥ 5 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 15.00% |
| | ≥ 9 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| BASC-2 Parent | | | | | | | | | |
| F Index | ≥ 1 | .70 | .70 | 2.30 | .43 | .67 | .73 | 30.43% | 70.00% |
| | ≥ 2 | .40 | .83 | 2.30 | .73 | .67 | .61 | 17.39% | 70.00% |
| | ≥ 3 | .40 | .83 | 2.40 | .72 | .67 | .63 | 13.04% | 40.00% |
| | **≥ 4** | **.35** | **.96** | **8.05** | **.68** | **.88** | **.63** | **4.35%** | **35.00%** |
| | ≥ 5 | .20 | .96 | 4.60 | .84 | .80 | .58 | 4.35% | 20.00% |
| | ≥ 6 | .10 | .96 | 2.30 | .94 | .67 | .55 | 4.35% | 10.00% |
| | ≥ 9 | .10 | 1.00 | 0.00 | .90 | 1.00 | .56 | 0.00% | 10.00% |
| | ≥ 12 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |
| BRIEF - Parent | | | | | | | | | |
| Negativity | ≥ 1 | .75 | .39 | 1.23 | .64 | .52 | .64 | 60.87% | 75.00% |
| | ≥ 2 | .60 | .52 | 1.25 | .77 | .52 | .60 | 47.83% | 60.00% |
| | ≥ 3 | .60 | .74 | 2.30 | .54 | .67 | .68 | 26.09% | 55.00% |
| | ≥ 4 | .40 | .83 | 2.30 | .73 | .67 | .61 | 17.39% | 40.00% |
| | **≥ 5** | **.35** | **.96** | **8.05** | **.68** | **.88** | **.63** | **4.35%** | **35.00%** |
| | ≥ 6 | .30 | .96 | 6.90 | .73 | .86 | .61 | 4.35% | 30.00% |
| | ≥ 7 | .15 | 1.00 | 0.00 | .85 | 1.00 | .58 | 0.00% | 15.00% |
| | ≥ 9 | .05 | 1.00 | 0.00 | .95 | 1.00 | .55 | 0.00% | 5.00% |

*Note:* Sn = sensitivity, Sp = specificity, LR+ = positive likelihood ratio; LR- = negative likelihood ratio; PPP = positive predictive power; NPP = negative predictive power; BASC-2 = Behavior Assessment System for Children – Second Edition; BRIEF = Behavior Rating Inventory of Executive Function

Appendix V

BASC-2 Parent and Child/Adolescent Correlation Coefficients for Overlapping Clinical Scales

| | Parent / Child Reports ($n$ = 42) | Parent/ Adolescent Reports ($n$ = 24) |
|---|---|---|
| Clinical Scales | | |
| Atypicality | .19 | .56** |
| Anxiety | .25 | .33 |
| Depression | .14 | .22 |
| Somatization | — | .42* |
| Attention Problems | .41** | .45* |
| Hyperactivity | .30 | .39 |

*Note:* BASC-2 = Behavior Assessment System for Children – Second Edition; * = $p > .05$; ** = $p > .01$

Appendix W

Frequency of Failure for Participants with ADHD ($n$ = 60) at Previously Established Cut-Off Scores

| | Cut-off | ADHD |
|---|---|---|
| VSVT[a] | | |
| Invalid | | |
| Easy Correct | ≤ 7 | 0% (0) |
| Difficult Correct | ≤ 7 | 0% (0) |
| Total Correct | ≤ 17 | 0% (0) |
| Questionable | | |
| Easy Correct | 8 - 15 | 7% (4) |
| Difficult Correct | 8 - 15 | 27% (16) |
| Total Correct | 18 - 29 | 8% (5) |
| Valid | | |
| Easy Correct | ≥ 16 | 93% (54) |
| Difficult Correct | ≥ 16 | 73% (44) |
| Total Correct | ≥ 30 | 92% (55) |
| Digit Span | | |
| ACSS[b] | ≤ 5 | 19% (11) |
| RDS[b] | ≤ 6 | 22% (13) |
| RDS-R[c] | ≤ 11 | 40% (21) |
| AST[e] | | |
| Alphabet[1] | ≥ 8 | 25% (3) |
| Counting[1] | ≥ 6 | 17% (2) |
| Days of Week[1] | ≥ 4 | 8% (1) |
| Months[1] | ≥ 10 | 42% (5) |
| Total[1] | ≥ 27 | 25% (3) |

*Note:* Sample size varies slightly across measures, thus frequency and percentage are reported. Cut-off scores were identified from the following publications; VSVT[a] = Slick et al., 1997 VSVT manual with adult normative data; ACSS[b] and RDS[b] = Kirkwood et al., 2011 child and adolescent mTBI sample; RDS-R[c] = Spencer et al. 2013 and Young et al., 2012 adult veterans; AST[e] = Kirkwood et al., 2014 child and adolescent mTBI sample; [1] = time in seconds

Appendix X

Frequency of PVT Failure Rate for Participants with ADHD ($n$ = 60) at Newly Proposed Cut-Off Scores

| PVT | Cut-Score | ADHD |
|---|---|---|
| VSVT | | |
| Easy Correct | ≤ 17 | 11% (7) |
| Dif. Correct | ≤ 12 | 15% (9) |
| Total Correct | ≤ 31 | 17% (10) |
| Digit Span | | |
| DS ACSS | ≤ 4 | 8% (5) |
| RDS | ≤ 5 | 10% (6) |
| RDS-R | ≤ 8 | 7% (4) |
| AST | | |
| Alphabet | ≥ 19 | 8% (1) |
| Count to 20 | ≥ 8 | 8% (1) |
| Days of Week | ≥ 6 | 8% (1) |
| Months of Year | ≥ 22 | 8% (1) |
| Total Time | ≥ 52 | 8% (1) |

*Note:* Cut-off scores presented were identified in Table 10 as optimal for the entire clinical sample. Frequency and percentage of failure rate is presented due to varied sample size across measures.

Appendix Y

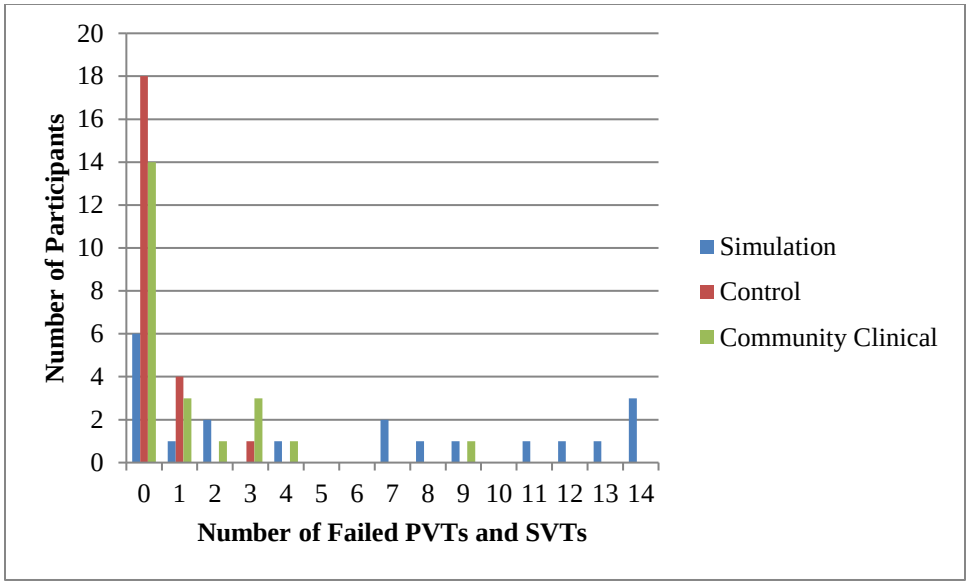


Figure 1. Number of PVTs and SVTs failed, organized by community- recruited group.