EMS PRIORITY 2-5 RESEARCH GRANT FINAL PROJECT REPORT

Agency Name: Ohio Department of Public Safety EMS Program

Project Title: Virtual Reality Cognitive Assessment Tool For Pediatric TBI

Reporting Period: 07/01/2018 – 06/30/2020

Table of Contents

ABSTRACT/EXECUTIVE SUMMARY

Background: The present study evaluated the psychometric reliability and validity of a novel VR-based Cognitive Assessment Tool (VR-CAT) as an alternative EF assessment tool for children with TBI. The VR-CAT focused on core executive functions following childhood TBI.

Methods: A total of 54 pediatric patients, including 30 children with orthopedic injuries and 24 children with traumatic brain injuries were recruited at Nationwide Children's Hospital, who were matched as close as possible to age and gender. A multi-method approach was adopted to evaluate the test-retest reliability, face validity, construct validity and diagnostic validity of the VR-CAT in comparison to gold-standard measures for EF assessment in the literature.

Results: The findings supported the face validity of all three EF games in the VR-CAT. The VR-CAT also showed adequate test-retest reliability, construct validity, and diagnostic validity for inhibitory control and working memory, but not cogniive flexibility skills.

Conclusions: The present study developed a promising protoype of an immersive assessment tool of exeuctive functions specifically designed for chilren with brain injuries and demonstrated comparable psychometric properties as those widely adopted in the literature.

QUALIFICATIONS OF PI AND CO-IS

The project was led by Dr. Jiabin Shen (PI), Research Scientist at Center for Biobehavioral Health at Nationwide Children's Hospital. Trained in developmental psychology, Dr. Shen has a strong interest and publication record in theory-driven and technology-assisted injury prevention, psychometric assessment development for children, and pediatric trauma research, including the use of virtual reality and randomized clinical trial designs for both TBI and burn pediatric patients. Dr. Shen was responsible for the overall design and implementation of this project, including protocol development, training of research staff, design and implementation of data analysis plan with center biostatistician, and research dissemination (manuscript writing, conference presentation, and translational media outreach).

Dr. Henry Xiang (Co-PI) is a senior scientist with a strong interest, funding and publication record in injury epidemiology and pediatric trauma research. He serves as the Director of the Center for Pediatric Trauma Research and the Director of Research Core of the Center for Injury Research and Policy at Nationwide Children's Hospital.

Dr. Christine Koterba (Co-Investigator) has extensive experience in pediatric neuropsychological assessment and is the attending neuropsychologist on the Nationwide Children's Hospital Inpatient Rehabilitation Unit. She has clinical and research interests in outcomes associated with pediatric acquired brain injury. She has worked closely with this research team on previous projects, including the pilot study which examined the use of a VR-based cognitive intervention on the Inpatient Rehabilitaiton unit.

Dr. Jeffrey Leonard (Co-Investigator) is Chief of Neurosurgery at Nationwide Children's Hospital and Professor of Neurological Surgery at The Ohio State University School of Medicine. Dr. Leonard has extensive clinical and research experience in children with traumatic

brain injuries with an in-depth understanding of how pediatric TBI could negatively influence children's short-term physical and cognitive abilities as well as long-term health outcomes.

Dr. Julie Samora (Co-Investigator) is a pediatric orthopaedic upper extremity surgeon with a background in basic science (microvascular physiology), public health, and healthcare administration. Dr. Samora serves as the Director of Upper Extremity Research and the Director of Quality Improvement at Nationwide Children's Hospital. She also currently serve on the American Academy of Orthopaedic Surgeons (AAOS) Committee on Evidence-Based Quality and Value, and previously served on the American Society for Surgery of the Hand (ASSH) Evidence Based Practice Committee, the Association of American Medical College's (AAMC) Advisory Panel on Health Care, and the Brigham and Women's Healthcare Safety Quality Improvement Committee (HSQC). Dr. Samora has completed an intensive year-long course on quality improvement essentials and developed a skill set to improve the children's experience in the orthopaedic clinic setting.

1. INTRODUCTION

Literature Review

The Centers for Disease Control and Prevention (CDC) classifies childhood traumatic brain injury (TBI) as the leading cause of death and acquired disability in children, with an estimated 700,000 childhood TBI cases every year in the United States (1-4). Defined as a disruption in the normal function of a child's brain that can be caused by a bump, blow, or jolt to the head, or a penetrating head injury, childhood TBIs often result in significant impairment in cognitive functions (1), particularly in executive functions (EFs) due to the vulnerability of the frontal lobes in children's developing brain (5-7). EF involves a set of core cognitive capacities for self-controlled discipline, creativity, and flexibility (8, 9). Deficits in EF have profound implications for the children's daily EF (8) and quality-of-life (QoL) (10), as reflected in increased attention problems (11), poorer academic performance (12), and poorer psychosocial adjustment (13). Therefore, fast and accurate assessment of EF impairment following childhood TBI is critical for to help emergency personnel and acute care providers design precise medical treatment and rehabilitation plans for pediatric patients with TBI.

However, evidence-based EF assessment tools specifically designed for efficiently screening EF deficits following childhood TBI are unavailable. Although a combination of computerized and non-computerized cognitive games has been developed for cognitive training among healthy children and pediatric patients (9,14), existing popular cognitive assessment tools, such as the NIH Toolbox Cognition Battery, have four key obstacles that hamper their successful implementation among children with TBI: affordability, accessibility, adherence, and generalizability. First, implementation of existing broad-spectrum EF assessment tools is both costly in time and medical expenses for families and resource-demanding for hospitals (15).

Second, existing EF assessment programs require the significant presence of both therapists and patients at the hospital or outpatient clinic, which is not only time- and resource-consuming, but also impedes accessibility for families living in remote areas, contributing to disparities in health outcomes. Third, existing EF assessment programs typically use rudimentary paper-and-pencil tasks or artistically substandard flat-screen computer tasks, both of which are unappealing to children and adolescents who are increasingly accustomed to engaging video games available to them outside the medical settings (16). This may negatively affect patients' willingness to participate in cognitive assessment activities, especially for repeated assessment during important follow-up sessions. Fourth, partially due to the limited scientific generalizability, existing EF assessment programs are often unable to fully identify all of the EF deficits in children that reflect real-life psychobehavioral limitations.

Current Status of VR Applications in Cogntive Assessment for Pediatric TBI

In recent years, virtual reality (VR) technology has rapidly emerged as a promising alternative strategy to address the above-mentioned limitations of traditional assessment tools like the NIH Toolbox Cognition Battery. First, VR has the capability to offer a multitude of activities for evaluating children's cognitive functions within a versatile virtual environment. Such an enriched, immersive assessment environment offers potentially high scientific transferability of the VR assessment results to patients' real-life functions. Second, because all VR-based cognitive assessment occurs within a safely controlled, automated virtual environment, it takes minimal physical space and personnel resources to implement in a medical setting including emergency departments, rehabilitation units, and concussion clinics, hence reducing expenses for both the hospitals and the families. Third, VR-based cognitive assessment was found especially appealing to pediatric patients with TBI in our pilot testing, largely due to the youth population's increasing

exposure to electronic games, including immersive VR games in recent years. This factor is important as it will increase the adoptation rate of cognitive assessment tools especially in busy clinics and inpatient units. Finally, upon validation of the reliability, scientific validity, and diagnostic utility of this proposed VR system, it can be easily morphed onto online and mobile computing platforms such as internet browser-based VR or smartphone-based VR systems. This will allow cognitive assessment, especially follow-up assessments scheduled after discharge from hospitals and emergency departments, to be completed without requiring patients to leave home. This accessibility further increases the compliance with long-term cognitive funtion assessment and monitoring needs.

Applying VR technology specifically to the post-brain injury assessment of EF is lacking compared to VR's application in other medical scenarios, such as pain management. A recently published meta-analytic review paper on VR-based measures in neuropsychological assessment and found that VR-based cognitive assessment has become increasingly popular due to its high scientific validity (17). However, the systematic review only identified five published studies in the literature involving VR-based EF assessment for patients with brain injuries, most of which focused on the adult population (18-21) and stroke patients (18-20). That said, all of them reported high feasibility and utility in their target populations. Due to the significant differences in the development of brain and cognitive skills between children and adults, pediatric patients with TBI well deserve a dedicated VR-based EF assessment system. Yet, the literature review showed that only one study, published almost five years ago (in 2013), compared EF performance within a virtual supermarket for children with TBI to typically developing children (21). Specifically, this study asked children with TBI to shop for four items in a virtual supermarket (VMall), followed by a brief feedback questionnaire and a Zoo Map task for

dysexecutive syndrome. The study reported good usability among children. The mean shopping time and number of mistakes was found higher among children with TBI compared to healthy children. However, the VR system used by this study wasn't validated by credible psychometric indices including test-retest reliability, internal and external (scientific) validity, or diagnostic utility. While promising, these findings leave a significant gap for both clinical research and practice in the field of VR-based EF assessment for pediatric patients with TBI.

The Present Study

The present study was designed to fill this critical gap in the field by evaluating the psychometric reliability and validity of a novel VR-based Cognitive Assessment Tool (VR-CAT) as an alternative EF assessment tool for children with TBI. The VR-CAT focused on three core executive functions following childhood TBI based on Diamond's Framework of Executive Functions (8): inhibitory control (the ability to override a strong internal predisposition or external lure and do what is more appropriate or needed), working memory (the ability to hold and process information in mind as needed), and cognitive flexibility (the ability to adjust to changing environmental demands and think with different perspectives). Speific hypotheses tested in the persent study are as follows:

Hypothesis 1. Pediatric patients, patient families, and clinicians will report high usability for VR-CAT among pediatric patients with TBI.

Hypothesis 2. VR-CAT will demonstrate high correlations between VR tests (test-retest reliability), between VR-CAT and NIH ToolBox (construct validity), and between VR-CAT and BRIEF-2 (ecological validity).

Hypothesis 3. The VR-CAT has high accuracy (> 90%) in differentiating between pediatric TBI and pediatric orthopedic injury based on EF functioning.

2. METHODS

2.1 Participants

We recruited 24 children with TBI (7 to 17 years old) and 30 age-, gender-, and injury severity-matched children with OI at Nationwide Children's Hospital. Participants were initially identified based on a daily review of the electronic medical record at Nationwide Children's Hospital. Inclusion criteria for children in the TBI group are patients 7 to 17 years of age (inclusive) who: 1) were diagnosed with a) mild complicated TBI or b) moderate to severe TBI (ICD-9 code of 803, 850, 851, 854) when admitted; 2) Glasgow Coma Scale (GCS) ≥ 13 plus presence of a depressed skull fracture or trauma related intracranial abnormality (complicated mild TBI), GCS = 9-12 (moderate TBI, CDC/NIH definition), and GCS =3-8 (severe TBI, CDC/NIH definition); 3) household use of English as the primary language; and 4) currently score 28 or below on the Agitated Behavior Scale (ABS), indicating mild to no agitation. Inclusion criteria for the OI group are pediatric patients matched as close as possible to age, gender, and injury severity of the TBI group who: 1) sustained an orthopedic injury requiring at least an overnight hospitalization; 2) no fracture associated with loss of consciousness or indicators of possible brain injury such as facial injuries.

Exclusion criteria for both the TBI and OI groups are: 1) patients who have severe pre-injury physical/visual/cognitive impairments that prevent proper utilization of a VR-based assessment and valid administration of the study measures, including autism, developmental delay, severe language impairment or neurologic disorder prior to their injury (per medical records); 2) patients who are on droplet/contact isolation; 3) patients who have known seizure activities post-injury; 4) patients who sustained their injuries due to abuse; 5) patients who have not been cleared from narcotics for pain, which might impair them cognitively.

2.2 Description of the VR-based Cognitive Assessment Tool (VR-CAT) for children with TBI.

The hardware of the VR-CAT system consists of a fully-immersive High Tech Computer (HTC) VIVE VR viewer (OLED display with 2160×1200 resolution, 90Hz refresh rate, and 110 degrees of field of view) powered by a high-performance laptop (Dell AlienwareTM 17), a pair of surround sound headphones reworked to integrate with the VR viewer, and a wireless one-hand controller for the child to interact with the virtual environment using gestures. Furthermore, to avoid the burden of the equipment's weight on a child's head, we created a special mechanical arm to suspend the VR viewer in front of a child's eyes with adjustable headphone positions. With the system securely fixed on a wheeled workstation, a child can operate the program by reclining on the bed/wheelchair in his/her own patient room. To ensure hygiene, each child will use a new facial pad for the viewer and a new pair of earphone covers. All parts of the equipment will be sanitized with germicidal wipes after each use.

The software of the VR-CAT system is a Windows 10-based virtual reality application that invites children to rescue an animated character named "Lubdub" from a heavily-guarded castle. The program consists of three child-friendly assessment tasks that correspond to the three core EFs. During the game, children need to (a) direct a group of sentinels away from the castle gates (VR Task #1, adapted from the Spatial Stroop Task to assess Inhibitory Control), (b) successfully open a series of castle gates by replicating the cryptography sequence of items surrounding each gate in forward/backward order (VR Task #2, adapted from the Visual Working Memory Task to assess Working Memory), and finally (c) rescue Lubdub inside the castle by strategically matching patterns between Lubdub and the four surrounding guards (VR Task #3, adapted from the Wisconsin Card Sorting Task to assess Cognitive Flexibility). A tutorial mode preceding each assessment task allows children to learn how to interact with VR

and ask researchers any questions they may have. The assessment mode consists of 30 trials per task and takes only 10 minutes in total to complete the entire assessment. Children's performance, including error rate and response time, will be automatically recorded by the VR-CAT computer system throughout the assessment.

2.3. Measures

Usability: A User Feedback Survey collected user feedback of the VR-CAT program, including level of discomfort using a 15-item Simulator Sickness Questionnaire and level of fatigability after intervention with the Borg Rating of Perceived Exertion Scale. This survey will be given to both TBI and OI patients after their initial experience with VR-CAT.

NIH Toolbox Cognition Battery: The NIH Toolbox Cognition Battery is a comprehensive set of standardized neuro-behavioral measurements used to assess individuals' cognitive functions using an iPad-based application for all children above seven years of age. Three EF-related tests within the Cognition Battery were implemented in this study: the Flanker Inhibitory Control and Attention Test (Inhibitory Control), List Sorting Working Memory Test (Working Memory), and Dimensional Change Card Sort Test (Cognitive Flexibility). This assessment takes 15 minutes to complete.

Behavior Rating Inventory of Executive Function-2nd Edition (BRIEF-2): The parent-reported BRIEF-2 questionnaire measures children's EF in daily environments such as home and school and offers us the opportunity to examine how scores obtained from the VR-CAT can 'ecologically' generalize to real-life scenarios. The BREIF-2 is completed by parents/legal guardians and takes 10 minutes to administer.

Demographic and Injury-Related Information: Patients' medical records were accessed to obtain the following demographic and injury-related variables of children: age, gender, and injury

severity score (ISS).

2.4 Procedures

The protocol of this study was approved by Nationwide Children's Hospital IRB (IRB16-00664) with detailed study procedures outlined below:

Screening and Consenting: Our research staff monitored the electronic medical record each day at Nationwide Children's Hospital and identify all potential patients who were admitted with a TBI or OI and who meet the inclusion/exclusion criteria. They then worked with our medical collaborators in the Nationwide Children's Hospital to identify and approach potential participants with TBI/OI. The Nationwide Children's Hospital trauma registry and admission records were also used to assist in identifying children who have been hospitalized for OI/TBI and then, using inclusion criteria, study eligibility were determined. Once deemed eligible (using inclusion/exclusion criteria) for the study, participants and their parents were approached during their hospital stay by a research staff member to introduce the research project. The staff member would wait until patients are alert, awake, and demonstrate understanding of the study by his/herself before approaching families. If the patient is ≥9 years of age, he/she must be able to sign the assent form. The parent (legal guardian) will also be consented for study participation.

Assessment #1: Following recruitment, participants were scheduled to complete Assessment #1, which includes: a) 10-minute VR-CAT followed by a 5-minute VR Experience Survey and b) 10-minute NIH Toolbox Cognition Battery tests. The order of tests within the Assessment #1 were randomized for each participant to minimize the order effect of consecutive testing. Tutorial training were provided to both VR-CAT tasks and NIH Toolbox tasks prior to test administration. Parents (or legal guardians) completed the BRIEF-2 while their child completes Assessment #1. Assessment #1 lasted about 30 minutes (including parent reports and tutorial

training). Depending on a patient's diagnosis and length of hospital admission, Assessment #1 were either administered in the hospital or in a follow-up outpatient clinic.

Assessment #2: Assessment #2 took place approximately one week following the administration of Assessment #1. Assessment #2 was designed to evaluate the test-retest reliability of VR-CAT and therefore only the 10-minute VR-CAT were included in this set of assessment and only for OI patients. The rationale for excluding the TBI group on Assessment #2 was that children with TBI have a natural recovery of brain functions, potentially rapidly, during the one week between Assessment #1 and #2; therefore they were not suitable for providing test-retest reliability data as the brain recovery might confound the accuracy of reliability assessment. On the other hand, the OI group did not have this concern. Due to our strict inclusion/exclusion and matching criteria for the OI group, children with OI were not expected to have significant brain functional changes over the course of one week, hence providing unbiased data for reliability assessment. Assessment #2 usually took place at the OI clinic.

2.5 Analytic Plan:

- Demograhics: Descriptive analysis was applied to demonstrate the demographics of the sample. Categorical demographics were reported as numbers and percentages. Numeric demographics (age, GCS) and all assessment scores at each timepoint were shown as mean and standard deviation (SD).

- Face Validity (Hypothesis 1): Usability of VR-CAT will be evaluated by calculating mean simulator sickness scores and mean Borg exertion scores and compared between TBI and OI groups using Wilcoxon rank-sum tests.

- Reliability (Hypothesis 2): Test-Retest Reliability (temporal stability) is a standard psychometric index to measure a neuropsycholgoical asessment's temporal stability. Specific to

this study, test-retest reliability will be computed as the Pearson correlation between OI patients' performance on the initial VR-CAT assessment and that on the second VR-CAT assessment one week later.

- Construct Validity (Hypothesis 2): Construct validity is defined as the degree to which a neuropsychological test measures what it intends to measure. Construct validity is usually evaluated by establishing the correlation between the test under evaluation and a 'standard' test commonly accepted in the field that measures the same construct. In the field of EF assessment, the NIH Toolbox Cognitive Battery consists of three standard neuropsychological tests on executive functions that have been widely adopted: the Flanker Inhibitory Control and Attention Test (Inhibitory Control), List Sorting Working Memory Test (Working Memory), and Dimensional Change Card Sort Test (Cognitive Flexibility). Thus, we assessed the VR-CAT's construct validity by examining how well pediatric patients' performance on the three VR-CAT tasks respectively correlated with the three above-mentioned standard NIH Toolbox tests using Pearson correlation analysis. We also examined the Pearson correlation of the total performance scores between the three VR-CAT tasks and the three NIH Toolbox tests scores.

- Ecological Validity (Hypothesis 2): The ecological validity of VR-CAT were computed by calculating the Pearson correlation between participants' performance on VR-CAT and their parents' or legal guardians' report on BRIEF-2 regarding their child's real-life EF.

- Diagnostic Utility (Hypothesis 3): Receiver Operating Characteristic (ROC) curves were plotted to compare the diagnostic ability among the three tools: VR-CAT, NIH toolbox and BRIEF2. We also compared the performance scores between TBI and OI groups on the same three measures as an alternative method to evaluate VR-CAT’s diagnostic utility.

3. RESULTS

Of the 54 participants who were matched in the cohort study, 30 children had orthopedic injuries (defined as OI group): 70% boys, half non-Hispanic Caucasian and averaged 12 years old; 24 children were with traumatic brain injury (defined as TBI group): 63% boys, more than 80% non-Hispanic Caucasian, averaged 13 years old, one third complicated mild cases and less than a half severe cases (Table 1).

Table 1. Description of Study Sample, by Injury Type (N=54)

Demographics	OI (N=30)	TBI (N=24)
Gender (n, %)		
Male	21 (70.00)	15 (62.50)
Female	9 (30.00)	9 (37.50)
Age (mean, SD)	11.70 (3.12)	12.75 (3.29)
Race (n, %)		
Caucasian/Non-Hispanic	15 (50.00)	20 (83.33)
African American/Non-Hispanic	10 (33.33)	3 (12.50)
Hispanic	1 (3.33)	0 (0.00)
Other	4 (13.33)	1 (4.17)
TBI-Only Characteristics		
GCS (mean, SD)	--	9.17 (4.74)
TBI Severity (n, %)		
Complicated Mild	--	8 (33.33)
Moderate	--	4 (16.67)
Severe	--	11 (45.83)

Table 2 shows the face validity of the patients' subjective experiences with VR-CAT at the first assessment. Overall, both OI and TBI groups of children scored relatively low on simulator sickness (mean=0.11, SD=0.14 and mean=0.05, SD=0.08, respectively) and physical exertion (mean=8.23, SD=2.36 and mean=8.79, SD=3.53, respectively). Most children felt like they were inside the game (mean=3.50, SD=1.27), had fun during the games (mean=4.20, SD=0.90), liked the VR games they played (mean=4.17, SD=0.99), would like to play again in

the future (mean=3.96, SD=1.18), would like to have the VR therapies in the hospital (mean=4.17, SD=1.15), and had motivation of attending this VR therapy if possible (mean=4.15, SD=1.09).

Table 2. Face Validity: Patients' Subjective Experiences with VR-CAT

Variables	Groups			
	Overall (N=54)	**OI (N=30)**	**TBI (N=24)**	***p***
Simulator Sickness (0-3)	0.08 (0.12)	0.11 (0.14)	0.05 (0.08)	0.09
Physical Exertion (6-20)	8.48 (2.92)	8.23 (2.36)	8.79 (3.53)	0.48
Immersive Virtual Environment (1-5)	3.50 (1.27)	3.73 (1.11)	3.21 (1.41)	0.13
Had Fun (1-5)	4.20 (0.90)	4.17 (0.95)	4.25 (0.85)	0.73
Liked the VR Games (1-5)	4.17 (0.99)	4.23 (0.97)	4.09 (1.04)	0.66
Wanted to Play Again in Future (1-5)	3.96 (1.18)	3.90 (1.27)	4.04 (1.08)	0.66
Wanted to See VR in Hospitals (1-5)	4.17 (1.15)	4.30 (0.92)	4.00 (1.38)	0.33
Motivated to Return to Clinical Appointments (1-5)	4.15 (1.09)	4.37 (0.81)	3.88 (1.33)	0.09

Table 3 presents the test-retest reliability for the VR-CAT instrument. Overall, the VR-CAT demonstrated adequate test-retest reliability ($r=0.55$, $p<0.01$). Specifically, inhibitory control ($r=0.66$, $p<0.01$) and working memory ($r=0.79$, $p<0.0001$) were highly correlated between the two assessments among patients with OI. However, cognitive flexibility was not significant correlated across the two assessment sets, although it still showed a positive correlational trend ($r=0.10$, $p=0.63$)

Table 3. Test-Retest Reliability of VR-CAT in Children with OI (N=30)

Variable	OI (N=30)	
	r	*p*
Inhibitory Control	0.66	<0.01
Working Memory	0.79	<0.0001
Cognitive Flexibility	0.10	0.63
Total	0.55	<0.01

Similar results were obtained when computing the construct validity, which was the correlation between VR-CAT and NIH toolbox (Table 4). Overall, the VR-CAT showed adequate construct validity among children with OI (r=0.52, p<0.01) and the combined sample of TBI and OI (r=0.49, p<0.01), but not the TBI only group although still a positive trend (r=0.32, p=0.13). Specific to each EF task, inhibitory control showed adequate construct validity with significant correlation between VR-CAT and NIH Toolbox tasks for combined sample (r=0.49, p<0.01), the OI group (r=0.40, p=0.03), and the TBI group (r=0.47, p=0.02). Working memory also showed adequate construct validity with significant correlation between VR-CAT and NIH Toolbox tasks for combined sample (r=0.49, p<0.01), the OI group (r=0.44, p=0.01), and the TBI group (r=0.50, p=0.02). However, cognitive flexibility is not significantly correlated in any group between VR-CAT and the NIH Toolbox task (overall: r=0.04, p=0.79; OI: r=0.15, p=0.42; TBI: r=-0.09, p=0.68).

Table 4. Concurrent Validity: Correlation Between VR-CAT and NIH Toolbox

Variable	Overall (N=54)		OI (N=30)		TBI (N=24)	
	r	*p*	*r*	*p*	*r*	*p*
Inhibitory Control	0.49	<0.01	0.41	0.03	0.47	0.02
Working Memory	0.49	<0.01	0.44	0.01	0.50	0.02
Cognitive Flexibility	0.04	0.79	0.15	0.42	-0.09	0.68
Total	0.49	<0.01	0.52	<0.01	0.32	0.13

Table 5 exhibits the ecological validity-correlation between VR-CAT and BRIEF2 questionnaire, which measured the executive function behaviors. In general, children's scores are positively correlated between VR-CAT and BRIEF2 Global Execution Composite (GEC) in the TBI group (r=0.57, p<0.01). Conversely, children in the OI group scored moderate and negative correlation between VR-CAT and GEC of BRIEF2 (r=-0.47, p=0.045). Also, in some sub-scales like Inhibit (OI: r=-0.41, p=0.02; TBI: r=0.45, p=0.03), Behavioral Regulation Index (BRI, OI:

r=-0.38, p=0.04; TBI: r=0.50, p=0.01), Task-Monitor (OI: r=-0.42, p=0.02; TBI: r=0. 47, p=0.02), and Cognitive Regulation Index (CRI, OI: r=-0.38, p=0.04; TBI: r=0.49, p=0.01), children's behaviors are negatively (for OI group) and positively (for TBI group) correlated between the two measures.

Table 5. Ecological Validity: Correlation Between VR-CAT and Parent-Reported BRIEF2 Scores

BRIEF2 T Scores	Overall (N=54)		OI (N=30)		TBI (N=24)	
	r	*p*	*r*	*p*	*r*	*p*
Inhibit	-0.06	0.67	-0.41	0.02	0.45	0.03
Self-Monitor	0.02	0.87	-0.25	0.19	0.51	0.01
Behavioral Regulation Index (BRI)	-0.04	0.76	-0.38	0.04	0.50	0.01
Shift	0.14	0.30	-0.17	0.37	0.55	0.01
Emotional Control	0.06	0.64	-0.23	0.23	0.52	0.01
Emotional Regulation Index (ERI)	0.11	0.44	-0.25	0.19	0.58	0.00
Initiate	0.02	0.89	-0.28	0.14	0.46	0.02
Working Memory	-0.07	0.63	-0.35	0.06	0.36	0.09
Plan/Organize	0.05	0.71	-0.33	0.07	0.52	0.01
Task-Monitor	-0.06	0.69	-0.42	0.02	0.47	0.02
Organization of Materials	0.05	0.73	-0.33	0.08	0.51	0.01
Cognitive Regulation Index (CRI)	-0.01	0.92	-0.38	0.04	0.49	0.01
Global Execution Composite (GEC)	0.02	0.89	-0.47	<0.05	0.57	<0.01

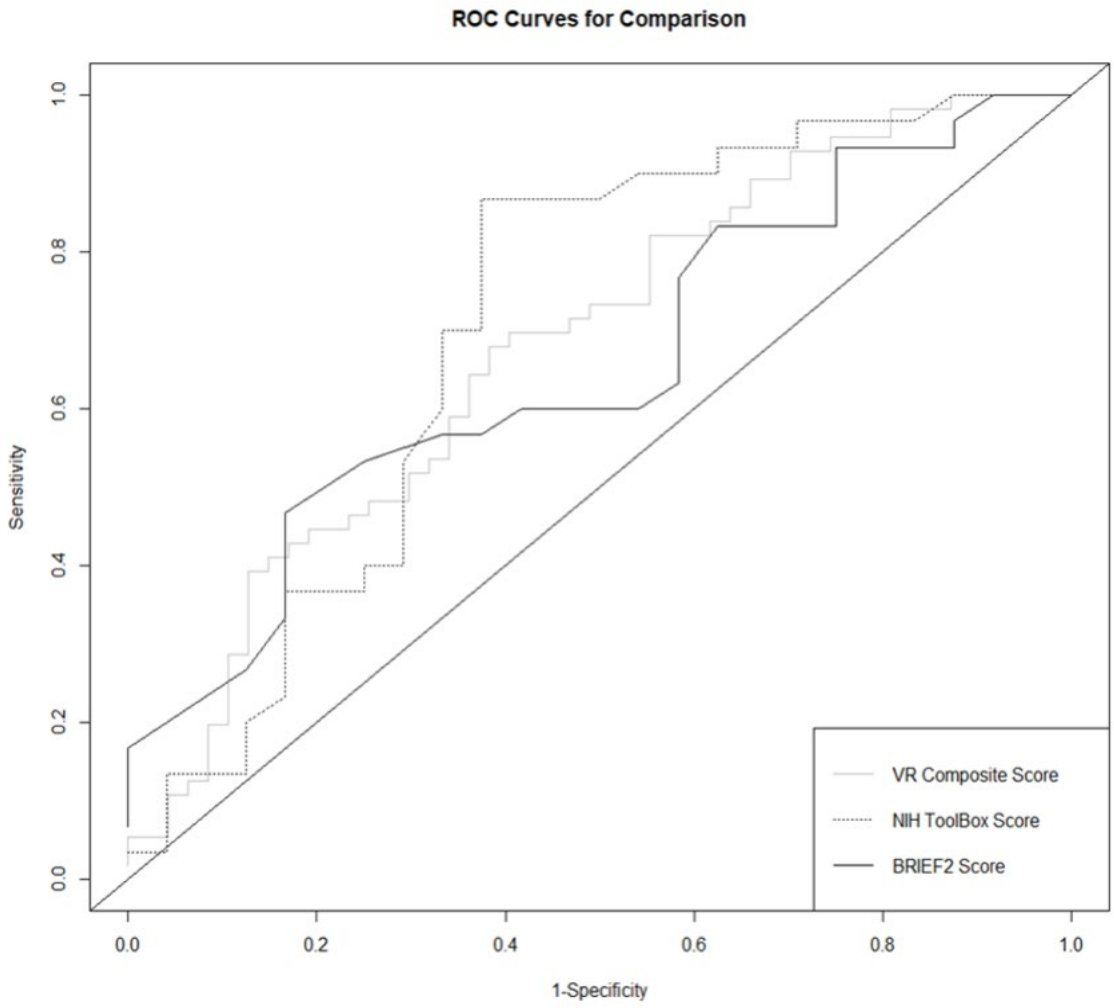
Table 6 shows the diagnostic validity-distinguishing behavioral results between TBI and OI groups on VR-CAT scores, NIH toolbox age-corrected scores, and BRIEF2 t-scores. Generally, children in OI groups showed higher EF performance than those in TBI groups on VR-CAT Composite Scores (OI: mean=102.99, SD=8.22, TBI: mean=95.11, SD=10.50, p<0.01), suggesting adequate ability of VR-CAT to distinguish between children with and without TBI diagnosis. Along the same line, it should also be mentioned that NIH Toolbox Composite Scores (OI: mean=100.23, SD=11.12, TBI: mean=90.61, SD=14.13, p<0.01), but not

any of the subscale or composite scores from BRIEF2, showed similar discriminant ability between TBI and non-TBI samples.

Table 6. Diagnostic Validity: Distinguishing Between TBI and OI Based on VR Scores, Compared to NIH Toolbox and BRIEF2

Instruments	OI (N=30)	TBI (N=24)	*p*
VR-CAT Scores			
VR Inhibitory Control	104.25 (12.67)	93.31 (17.27)	0.01
VR Working Memory	101.55 (13.97)	88.82 (11.70)	<0.01
VR Cognitive Flexibility	103.16 (15.24)	102.34 (15.15)	0.84
VR Composite Score	102.99 (8.22)	95.11 (10.50)	<0.01
NIH Toolbox Age-Corrected Scores			
Flanker Inhibitory Control and Attention Test	92.47 (14.89)	84.21 (16.67)	0.06
List Sorting Working Memory Test	113.77 (18.57)	99.96 (17.48)	0.01
Dimensional Change Card Sort Test	94.47 (13.56)	87.67 (17.02)	0.11
NIH Toolbox Composite Score	100.23 (11.12)	90.61 (14.13)	<0.01
BRIEF2 T Scores			
Inhibit	49.50 (10.93)	54.08 (11.01)	0.13
Self-Monitor	48.63 (10.70)	53.63 (10.37)	0.09
Behavioral Regulation Index (BRI)	48.97 (10.83)	54.33 (10.78)	0.08
Shift	49.13 (9.69)	51.71 (11.00)	0.37
Emotional Control	49.13 (8.80)	53.67 (10.13)	0.08
Emotional Regulation Index (ERI)	49.10 (8.41)	52.96 (10.32)	0.14
Initiate	47.90 (9.06)	53.29 (11.48)	0.06
Working Memory	50.83 (11.08)	55.29 (11.35)	0.15
Plan/Organize	48.80 (9.44)	53.83 (12.64)	0.10
Task-Monitor	48.90 (9.01)	54.50 (10.50)	0.04
Organization of Materials	49.93 (9.06)	51.83 (9.15)	0.45
Cognitive Regulation Index (CRI)	49.43 (10.31)	54.42 (11.59)	0.10
Global Execution Composite (GEC)	49.27 (10.27)	54.79 (11.65)	0.07

The ROC curves further compared the diagnostic validity among the three measures: VR-CAT, NIH toolbox and BRIEF2 (Figure 1). The area under the ROC curves (AUC) showed that VR-CAT is a more accurate tool than BRIEF2, but less than NIH toolbox (AUC (95% confidence interval) = 0.68 (0.58, 0.78); 0.65 (0.51, 0.80); 0.71 (0.56, 0.86) ; respectively).



4. DISCUSSION AND IMPLICATIONS

The present study validated the reliability and validity of a novel assessment tool of executive functions for children with TBI. It aligned with the mission of The Emergency Medical, Fire, & Transportation Services (EMFTS) Board as well as its Priority 2: "*Consideration will be given to entities that research, test, and evaluate medical procedures and systems related to adult and pediatric trauma care*". The VR-CAT is the first scientifically-validated VR-based EF assessment tool specifically designed for pediataric TBI patients, which received positive user

feedback not only from children, but also parents, trauma surgeons, and neuropsychologists. This unique VR cognitive assessment system consists of novel and developmentally-appropriate assessment content, fully immersive VR for high ecological validity, and a custom-designed mechanical support system to accommodate the clinical and physical needs of children with TBI. The innovative cognitive assessment tasks were developed based on cognitive science and neuropsychology with special consideration of the developmental and clinical characteristics of children with TBI.

Despite the increasing use of VR in medical settings, the present study was among the first to apply age-, gender-, and injury severity-matched case-control design to rigorously evaluate the reliability, validity, and diagnostic utility of a VR-based EF assessment tool for childhood TBI. The study also used multiple methods to assess not only the test-retest reliability (temporal stability) of VR-CAT, but also the multi-dimension validity of VR-CAT, including how performance on VR-CAT is associated with both standardized EF measures such as NIH Toolbox Cognition Battery (construct validity), as well as its association with real-life EF performance using parent-report BRIEF-2 behavioral questionnaire (ecological validity). This provided substantive evidence for the generalizability of VR-CAT, which has been lacking among existing cognitive assessment tools such as the NIH Toolbox.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The present study provided critical empirical evidence in developing this prototype of VR-CAT into a full-blown and population norm-based cognitive assessment system that can be utilized by medical providers for fast and accurate screening of cognitive deficits among pediatric patients with TBI. In future research, the VR-CAT can be readily morphed into internet browser-based VR or smartphone-based VR systems, allowing critical cognitive assessment, especially those

scheduled after discharge, to be completed remotely. This accessibility further increases the compliance with long-term cognitive funtion assessment and monitoring needs.

6. REFERENCES

1. Faul M, Xu L, Wald MM, Coronado V. Traumatic brain injury in the United States. Atlanta, GA: National Center for injury Prevention and Control, Centers for Disease Control and Prevention. 2010.
2. Centers for Disease Control and Prevention. Report to congress on traumatic brain injury in the United States: epidemiology and rehabilitation. Atlanta, GA: National Center for Injury Prevention and Control. 2014:1-72.
3. Keenan HT, Bratton SL. Epidemiology and outcomes of pediatric traumatic brain injury. Developmental Neuroscience. 2006;28(4-5):256-263.
4. Hiu Lam W, Mackersie A. Paediatric head injury: incidence, aetiology and management. Pediatric Anesthesia. 1999;9(5):377-385.
5. Scheibel RS, Levin HS. Frontal lobe dysfunction following closed head injury in children: Findings from neuropsychology and brain imaging. Development of the Prefrontal Cortex: Evolution, Neurobiology, and Behavior. 1997: 241-260.
6. Scott JG, Schoenberg MR. Frontal lobe/executive functioning. In: The Little Black Book of Neuropsychology. Springer. 2011:219-248.
7. Leblanc N, Chen S, Swank PR, Ewing-Cobbs L, Barnes M, Dennis M, et al. Response inhibition after traumatic brain injury (TBI) in children: Impairment and recovery. Developmental Neuropsychology. 2005;28(3):829-848.
8. Diamond A. Executive functions. Annual Review of Psychology. 2013;64:135-168.
9. Diamond A, Lee K. Interventions shown to aid executive function development in

children 4 to 12 years old. Science. 2011;333(6045):959-964.

10. Mangeot S, Armstrong K, Colvin AN, Yeates KO, Taylor HG. Long-term executive function deficits in children with traumatic brain injuries: Assessment using the Behavior Rating Inventory of Executive Function (BRIEF). Child Neuropsychology. 2002;8(4):271-284.

11. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biological Psychiatry. 2005;57(11):1336-1346.

12. Diamantopoulou S, Rydell A-M, Thorell LB, Bohlin G. Impact of executive functioning and symptoms of attention deficit hyperactivity disorder on children's peer relations and school performance. Developmental Neuropsychology. 2007;32(1):521-542.

13. Riggs NR, Blair CB, Greenberg MT. Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. Child Neuropsychology. 2004;9(4):267-276.

14. Carruthers K, Zampieri C, Damiano D. Relating motor and cognitive interventions in animals and humans. Translational Neuroscience. 2014;5(4):227-238.

15. Brooks C, Lindstrom J, McCray J, Whiteneck GG. Cost of medical care for a population-based sample of persons surviving traumatic brain injury. Journal of Head Trauma Rehabilitation. 1995;10(4):1-13.

16. Greenfield PM. Mind and Media: The Effects of Television, Video Games, and Computers. Psychology Press. 2014.

17. Neguț A, Matu SA, Sava FA, David D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. The Clinical Neuropsychologist. 2016 Feb

17;30(2):165-84.

18. Kang YJ, Ku J, Han K, Kim SI, Yu TW, Lee JH, Park CI. Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. CyberPsychology & Behavior. 2008 Jun 1;11(3):329-39.

19. Rand D, Rukan SB, Weiss PL, Katz N. Validation of the Virtual MET as an assessment tool for executive functions. Neuropsychological rehabilitation. 2009 Aug 1;19(4):583-602.

20. Rand D, Katz N, Weiss PL. Evaluation of virtual shopping in the VMall: Comparison of post-stroke participants to healthy control groups. Disability and rehabilitation. 2007 Jan 1;29(22):1710-9.

21. Pugnetti L, Mendozzi L, Attree EA, Barbieri E, Brooks BM, Cazzullo CL, Motta A, Rose FD. Probing memory and executive functions with virtual reality: Past and present studies. CyberPsychology & Behavior. 1998;1(2):151-61.

22. Erez N, Weiss PL, Kizony R, Rand D. Comparing performance within a virtual supermarket of children with traumatic brain injury to typically developing children: A pilot study. OTJR: occupation, participation and health. 2013 Oct;33(4):218-27.