Performance validity testing via telehealth and failure rate in veterans with moderate-to-severe traumatic brain injury: A veterans affairs TBI model systems study[1]

Robert J. Kanser[a], Justin J.F. O'Rourke[b] and Marc A. Silva[a,c,d,e,*]

[a]*Mental Health & Behavioral Sciences Section (MHBSS), James A. Haley Veterans' Hospital, Tampa, FL, USA*
[b]*Polytrauma Section, Audie L. Murphy Memorial Veterans' Hospital, San Antonio, TX, USA*
[c]*Department of Internal Medicine, University of South Florida, Tampa, FL, USA*
[d]*Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa, FL, USA*
[e]*Department of Psychology, University of South Florida, Tampa, FL, USA*

Abstract.

BACKGROUND: The COVID-19 pandemic has led to increased utilization of teleneuropsychology (TeleNP) services. Unfortunately, investigations of performance validity tests (PVT) delivered via TeleNP are sparse.

OBJECTIVE: The purpose of this study was to examine the specificity of the Reliable Digit Span (RDS) and 21-item test administered via telephone.

METHOD: Participants were 51 veterans with moderate-to-severe traumatic brain injury (TBI). All participants completed the RDS and 21-item test in the context of a larger TeleNP battery. Specificity rates were examined across multiple cutoffs for both PVTs.

RESULTS: Consistent with research employing traditional face-to-face neuropsychological evaluations, both PVTs maintained adequate specificity (i.e., > 90%) across previously established cutoffs. Specifically, defining performance invalidity as RDS < 7 or 21-item test forced choice total correct < 11 led to < 10% false positive classification errors.

CONCLUSIONS: Findings add to the limited body of research examining and provide preliminary support for the use of the RDS and 21-item test in TeleNP via telephone. Both measures maintained adequate specificity in veterans with moderate-to-severe TBI. Future investigations including clinical or experimental "feigners" in a counter-balanced cross-over design (i.e., face-to-face vs. TeleNP) are recommended.

Keywords: Brain injuries, traumatic, neuropsychological tests, performance validity, telemedicine, TBI

1. Introduction

Within neurorehabilitation, neuropsychological evaluations play a critical role in the diagnosis, monitoring, and treatment of cognitive and neurobehavioral dysfunction following neurological injury or disease. The utility of neuropsychological evaluations rests on the assumption that examinees are

performing to the best of their ability. For example, neuropsychological functioning during the early recovery from traumatic brain injury (TBI) predicts important rehabilitation outcomes such as employment (Sherer et al., 2002), productivity (Green et al., 2008), and driving fitness (Egeto et al., 2019). However, accurate prognostication by clinical providers requires credible performance by examines during neuropsychological testing.

1.1. Performance validity

Unfortunately, there exist myriad factors that threaten the validity of neuropsychological evaluation results. The behavior of examinees as it relates to the credibility of neuropsychological test results is broadly referred to as *performance validity*; however, the terms response bias, effort, and engagement have also been used in the literature that has accumulated during the past 20 years (Lippa, 2018). Invalid performances can be driven by external gains (e.g., malingering to secure financial benefit or avoid responsibility), internal gains (e.g., factitious disorder), unconscious drives (e.g., somatization, conversion disorder), diminished neurogenic drive (e.g., abulia, akinesia), or simply impacted by sensory impairments (e.g., visual/auditory deficits). Importantly, performance invalidity across its forms is a prominent and prevalent concern, accounting for as much as 50% of the variability in neuropsychological test scores (Green, Rohling, Lees-Haley, & Allen, 2001). Historically a focus of forensic neuropsychologists, the investigation (and measurement) of performance invalidity is of clinical concern as well. Base rates of performance invalidity during neuropsychological assessment vary across settings (i.e., litigation, clinical/rehabilitation) and approaches 40% in the presence of external incentives or when referrals specify medical conditions commonly encountered in rehabilitation settings (e.g., mild TBI, chronic pain, neurotoxin injury; Mittenberg, Patton, Canyock, & Condit, 2002). Failure to recognize this prevalent problem in the rehabilitation setting results in the inaccurate interpretation of one's neurocognitive functioning, potentially leading to inappropriate decision-making.

Accordingly, considerable research has been dedicated to the development of performance validity tests (PVTs) which are objective measures designed to distinguish between valid and invalid neuropsychological test performance. Because of the prevalence and impact of performance invalidity, the inclusion of PVTs during neuropsychological evaluations has become a practice standard (Chafetz et al., 2015; Heilbronner et al., 2009; Sweet, Benson, Nelson, & Moberg, 2015).

1.2. Neuropsychological testing and telehealth

The COVID-19 pandemic brought about significant change to standard practice, as neuropsychologists across the nation (in rehabilitation hospitals as well as traditional outpatient settings) were forced to suspend traditional face-to-face neuropsychological evaluations (Bilder et al., 2020). Practitioners were faced with a challenging decision: suspend evaluation services entirely or provide service through alternative modalities such as telehealth which the field of neuropsychology has been slow to adopt. However, the provision of teleneuropsychology (TeleNP) services would provide a means to safely deliver (albeit in a limited form) services to persons who need of neuropsychological evaluations while maintaining travel restrictions and distancing recommendations per the Centers of Disease Control and Prevention vis-à-vis the coronavirus pandemic.

In the rehabilitation setting, the decision to halt services would risk delaying diagnoses of possible neurocognitive and neuropsychiatric disorders, initiation of interventions such as cognitive rehabilitation and evidence-based therapies for psychological disorders, and monitoring of recovery (e.g., following pharmacological treatment or neurosurgical interventions). The effects of delayed diagnosis and treatment have a substantial psychosocial and economic cost (Rosado et al., 2018; Glen et al., 2020; Donders 2019). In contrast, the decision to provide services through innovative modalities via TeleNP requires a mental shift from the practice standard including the way that neuropsychological tests were historically developed and standardized.

Neuropsychology practice has been criticized for failing to adopt telehealth and other technological advances embraced by other health care disciplines (Miller & Barr, 2017). Investigations into the feasibility, validity, and clinical utility of teleNP began emerging in the late 20th century, although this research was limited in scope, infrequently appeared in scientific journals, and failed to meaningfully impact the field. However, the coronavirus pandemic has forced the field of neuropsychology to prioritize TeleNP in both research and clinical practice. Indeed, in response to the global health crisis caused by the coronavirus pandemic, the Inter-

organizational Practice Committee (comprised of the American Academy of Clinical Neuropsychology/American Board of Clinical Neuropsychology, National Academy of Neuropsychology/Division 40 of the American Psychological Association, the American Board of Professional Neuropsychology, and the American Psychological Associates Services, Inc.) issued guidance on TeleNP practice (Bilder et al., 2020).

There now exists a growing body of research supporting the reliability and validity of neuropsychological assessment delivered remotely via telehealth (e.g., Brearly et al., 2017; Marra, Hamlet, Bauer, & Bowers, 2020). Notably, the majority of this evidence base utilized subjects who were tested under the controlled conditions of a remote clinic, oftentimes with the use of a technician to establish videoconference connection and/or assist with task administration (Bilder et al., 2020). Much of the burgeoning research on TeleNP has used web-conferencing platforms as the service delivery method; while innovative, concerns about web-conferencing neuropsychological service delivery include access, particularly among rural or economically disadvantaged persons (i.e., the digital divide; Ramsetty & Adams, 2020; Van Dijk, 2017). Even among persons with internet access, criticism of TeleNP has been raised based on research showing that reliability and validity of TeleNP are impacted by patient characteristics (e.g., age, comfort and familiarity with computers) and technology factors such as monitor size, screen resolution, and internet speed (Brearly et al., 2017; Hirko et al., 2020). These access barriers have led to deeper consideration of TeleNP via telephone (Caze et al., 2020; Lacritz et al., 2020). Indeed, TeleNP via telephone has been effectively utilized in research settings dating back to the 1980s and continuing to present day (Brandt et al., 1988; Bunker et al., 2017; Castanho, Amorim, Zihl, Palha, Sousa, Santos, 2014; Castanho et al., 2016; Matchanova et al., 2020; Nelson et al., 2020; Pendlebury, Welch, Cuthbertson, Mariz, Mehta, & Rothwell, 2013; Plassman, Newman, Welsh, & Helms, 1994).

1.3. TeleNP and PVTs

Regardless of neuropsychology service delivery method, including TeleNP, performance validity remains a primary concern. Unfortunately, research into the use of PVTs in TeleNP is sparse. A vital step in PVT development is establishing adequate specificity in groups with known neurocognitive impairment. In this context, specificity reflects the proportion of participants with valid performance accurately identified as such by the PVT [Specificity = True Negatives ÷ (True Negatives+False Positives)] with adequate specificity defined as values > .90 by convention. In contrast, sensitivity reflects the proportion of participants with invalid performance (e.g., poor effort or task engagement) accurately identified by the PVT [Sensitivity = True Positives ÷ (True Positives+False Negatives]). In clinical contexts, specificity is prioritized over sensitivity because false-positive errors (i.e., incorrectly labeling impaired cognitive performance as invalid) are considered more detrimental to the patient than false negative errors (i.e., incorrectly labeling invalid performance that appears to show cognitive impairment as valid). As with traditional in person PVT development, a critical first step in establishing PVT utility in TeleNP is to demonstrate a low rate of false positive classification errors among examinees, especially among those with bona fide cognitive impairment. Thankfully, there exist a number of candidate PVTs that are robust to the effects of genuine neurocognitive impairment, demonstrating excellent specificity across a range of clinical groups (Lippa 2018). Surprisingly, there is a paucity of research on PVTs administered via TeleNP using telephone-based methods of communication. As such, the primary objective of the current study was to report the specificity of two well-established PVTs administered via TeleNP in patients with traumatic brain injury (TBI) admitted for inpatient neurorehabilitation. We selected two PVTs (See Methods) that rely solely on verbal communication given the concerns noted above as well as the finding that verbally mediated neuropsychological tasks translate more reliability and validly to TeleNP than visually dependent tasks (Brearly et al., 2017). It was hypothesized that these PVTs would maintain adequate specificity, consistent with previously published research utilizing comparable clinical groups.

2. Method

2.1. Participants

This study was a secondary analysis of the Veterans Affairs TBI Model Systems (VA TBIMS) dataset. The VA TBIMS is a prospective longitudinal multicenter study of rehabilitation outcomes following inpatient TBI rehabilitation (Lamberty et al., 2014;
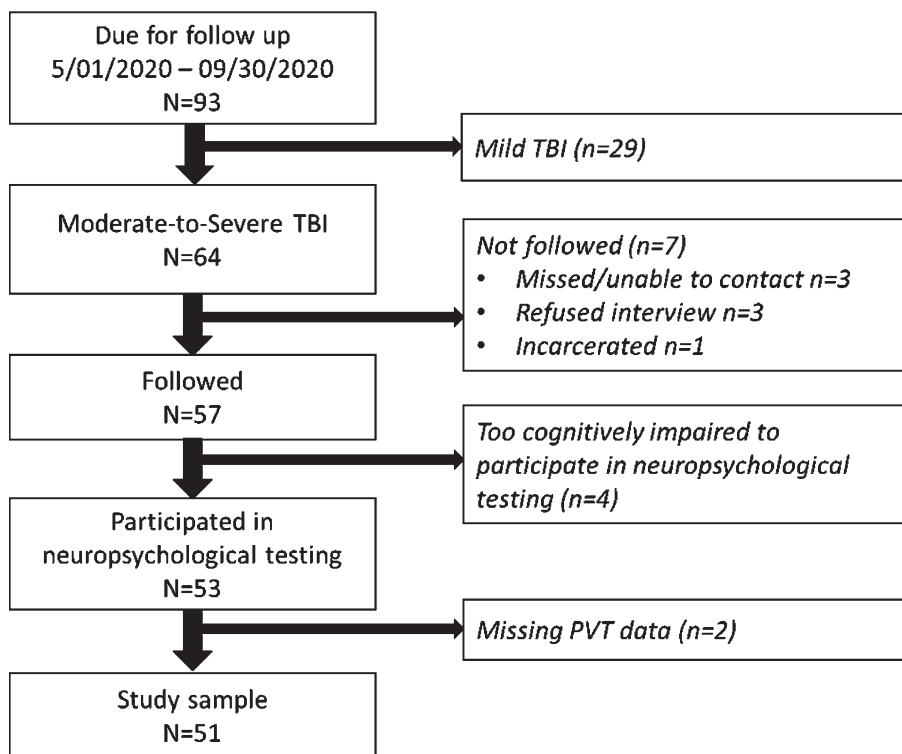
Fig. 1. Flow diagram.

Ropacki et al., 2018). This study included partici-pants enrolled at two study sites, the VA Polytrauma Rehabilitation Centers (PRCs) in San Antonio, Texas and Tampa, Florida. The CARF-accredited VA PRCs are rehabilitation programs that provide comprehen-sive inpatient rehabilitation services for veteran and active duty service members incurring injuries affect-ing the brain and/or other bodily systems. VA TBIMS inclusion criteria as follows: (1) TBI diagnosis per case definition: a self-reported or medically docu-mented traumatically-induced structural brain injury or physiological disruption of brain function due to external force evidenced by onset or worsening of: loss/decreased consciousness; mental state alteration; memory loss for events immediately before or after the injury; transient or stable neurological deficits; or intracranial lesion; (2) age ≥ 16 years at time of TBI; (3) admission to a PRC for TBI rehabilitation; and (4) informed consent by the participant or legally autho-rized representative. Additional inclusion criteria for this secondary analysis was (5) data collected after March 01, 2020 (which corresponded with the time that performance validity testing was added to data collection at the San Antonio and Tampa PRCs), and (6) medically documented moderate or severe TBI

based on established metrics (see section 2.2.3 TBI Characterization, below). Excluded were participants not followed after March 01, 2020, or participants were too cognitively impaired to participate in neu-ropsychological testing, or who were missing one or more PVT metrics. See Fig. 1 for flow diagram. Study design and procedures were approved by each site's Institutional Review Board.

2.2. Measures

2.2.1. 21-item test

The 21-item test (Iverson, 1998) is a measure designed to detect performance invalidity with a com-monly used forced-choice paradigm. Examinees are read a list of 21 words and then asked to recall as many words as possible. Immediately afterward, they are read 21 pairs of words, with one word in the pair from the word list and the other word a foil; they are asked to choose which word from the word pair was from the list. Classification of performance validity is based upon the total number of correct responses during the forced choice portion of the test. The 21-item forced choice total correct (21-FC-TC) score has a possible range of 0 to 21. Cutoffs

of < 8 and < 11 have been proposed (Gontkovsky & Souheaver, 2000; Iverson, 1998), with a cut score of < 11 showing optimal group classification based on the performance of 175 controls, 231 patients with psychiatric, neuropsychological, or memory disorders, and 237 experimental-malingerers (Iverson, 1998).

2.2.2. Reliable digit span (RDS)

RDS is an embedded performance validity measure derived from performance on the Digit Span subtest, which in this study was taken from the Wechsler Adult Intelligence Scale, third edition (Weschler, 1997). The WAIS is currently in its fourth edition, yet majority of the evidence on RDS pertains to the WAIS-III version (e.g., Babikian, Boone, Lu, & Arnold, 2006), although item content between the versions are similar. Examinees are asked to repeat sequences of digits of increasingly longer digit strands, with two trials of digits of the same length. Examinees repeat the digits first in forward order (simple repetition) and then in backward order (reversal of digits). Simply stated, RDS is the number of digits that are correctly repeated across both trials of the same digit length for the forward and backward conditions. RDS has a possible score range from 0 to 17. Various cut scores have been examined across existing studies (typically no lower than < 5). The cutoffs of < 6 and < 7 are the most used and have the best specificity and sensitivity across studies (Babikian, Boone, Lu, & Arnold, 2006; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012).

2.2.3. TBI characterization

TBI severity was characterized using the Glasgow Coma Score (Teasdale et al., 2014) at time of emergency department visit or hospital admission, time to follow commands (TFC) and duration of post-traumatic amnesia (PTA). TFC was the time elapsed (in days) between TBI and command following on two consecutive evaluations; PTA duration was the number of days between TBI and resolution of disorientation to person, place and time. TFC and PTA were calculated from review of acute and rehabilitation medical records following standardized procedures (Lamberty et al., 2014). If discharged from rehabilitation while still in PTA, duration was calculated as length of stay plus 1 day (Nakase-Richardson et al., 2011; Ropacki et al., 2018). To be classified as having moderate or severe TBI, participants were required to have medical documentation of GCS score ≤ 12, or TFC ≥ 1 day, or PTA duration ≥ 1 day, consistent with prior studies (Dillahunt-Aspillaga et al., 2017; Ropacki et al., 2018).

2.3. Procedures

Demographic characteristics were collected via structured interview (Lamberty et al., 2014; Ropacki et al., 2018) upon enrollment into the VA TBIMS. Injury information (e.g., TBI severity, cause of injury) were collected via review of medical records from emergency departments and acute care hospitalizations. VA TBIMS procedures include quarterly audit of research records, in which a research team member other than the research assistant initially collecting the data reviews data collection forms and source materials (i.e., hospital records) to ensure accuracy of coded data. Any discrepancies that arise are reviewed by the team's research coordinator and coded data are corrected if necessary.

PVTs were administered alongside neuropsychological tests via telephone with participants following inpatient rehabilitation discharge at regularly scheduled follow up time points anchored to TBI anniversary date. PVT failure rates at selected cutoffs were presented as percentages. Demographic and injury characteristics were presented as means and standard deviations for continuous data and as percentages for categorical data.

3. Results

3.1. Participants

Data on PVT performance were collected between March 01, 2020 and September 30, 2020. Of 93 participants due for follow up in this time frame, 64 had medically documented moderate-to-severe TBI. Of these, 89% were followed. However, 7% of those followed were unable to participate in testing because they were too severely cognitive impaired (e.g., prolonged disorders of consciousness), and 4% of those who participated in testing were missing PVT data. This resulted in a final sample size of 51 participants, which was 96% of those with TeleNP neuropsychological testing during follow up (see flow diagram, Fig. 1). Participants were 100% male, with an average age of 34 years (SD = 14) at time of TBI, over 55% had more than a high school education, and 71% identified as White. Cause of injury was most often vehicular related (61%). See Table 1 for further details of study sample demographic and injury char-

<div style="float:left">

Table 1
Characteristics of study participants ($N = 51$)

	M $\pm$ SD or %
Age at time of TBI	34.3 $\pm$ 13.7
Age at time of cognitive testing	38.1 $\pm$ 14.1
Sex	
Male	100
Female	0
Race/ethnicity[1]	
Hispanic, yes	13.7
White, yes	70.6
Black, yes	21.6
Asian, yes	3.9
Native American, yes	9.8
Pacific Islander, yes	3.9
Education	
< 12 years	7.8
High school diploma	35.3
Some college	41.2
Bachelor's degree	11.8
> Bachelor's degree	3.9
Annual income at time of TBI	
$\leq$ 9999	5.9
10,000–19999	11.8
20,000–29999	19.6
30,000–39999	5.9
40,000–49999	5.9
50,000–59999	7.8
60,000–69999	7.8
70,000–79999	5.9
80,000–89999	0
90,000–99999	2.0
$\geq$ 10000	3.9
Not applicable (not employed)	17.6
Unknown	5.9
Injury cause	
Vehicular	60.8
Fall	17.6
Bicycle/pedestrian	9.8
Blast	0
Assault	3.9
Gunshot wound	3.9
Sports-related	0
Other	2.0
Active duty at time of injury, yes	54.9
Injured during deployment, yes	9.8
GCS categories ($n = 51$)	
3–8 (severe)	68.6
9–12 (moderate)	9.8
13–15 (mild)	19.6
No acute care	0
GCS continuous format ($n = 50$)	6.5 $\pm$ 4.4
TFC, days ($n = 47$)	10.5 $\pm$ 16.3
PTA duration, days ($n = 48$)	42.7 $\pm$ 47.1
Rehabilitation length of stay, days	52.7 $\pm$ 51.5
TSI at time of PVT testing, months	44.4 $\pm$ 25.0
21-item test forced choice total correct	16.4 $\pm$ 3.1
Reliable Digit Span score	10.5 $\pm$ 2.4

[1]Race ethnicities do not add to 100% because people may endorse multiple categories. GCS = Glasgow Coma Score; PTA = Posttraumatic Amnesia; PVT = Performance Validity Test(s); TFC = Time to Follow Commands; TSI = Time Since Injury.

</div>

Table 2
PVT failure rate at various cutoffs ($N = 51$)

	Cutoff for failure	Failure rate (%)
21-item test force choice total correct	≤ 8	2.0, N = 1
	≤ 9	2.0, N = 1
	≤ 10	3.9, N = 2
	≤ 11	7.8, N = 4
	≤ 12	13.7, N = 7
Reliable Digit Span	≤ 5	2.0, N = 1
	≤ 6	2.0, N = 1
	≤ 7	9.8, N = 5
	≤ 8	19.6, N = 10

Note: Failure rate is the proportion of the sample with scores at or below the specified raw score.

acteristics. TeleNP neuropsychological testing took place between 1 to 7 years post-TBI.

3.2. PVT performance

21-FC-TC scores ranged from 7 to 21 ($M = 16.4$; $SD = 3.1$). At four of the five cutoffs, failure rate was less than 10% with a range of 2.0% to 7.8%. At the liberal cutoff of < 12, failure rate was 13.7% (see Table 2). RDS scores ranged from 5 to 16 ($M = 10.5$; $SD = 2.4$). At three of the four cutoffs, failure rate was less than 10% with a range of 2.0% to 9.8%. However, a liberal cutoff of < 8 resulted in a failure rate of 19.6% (see Table 2).

4. Discussion

Findings provide preliminary support for the hypothesis that verbally mediated PVTs are viable assessment tools in TeleNP evaluations. The RDS and 21-item test demonstrated adequate specificity across several proposed cut-offs in individuals with moderate to severe TBI. Despite the numerous concerns surrounding TeleNP and the effects of genuine brain injury on PVT performance, the remote telephone-administration of the RDS and 21-item test did not lead to rates of invalid performance that exceed acceptable specificity rates among individuals with known neurocognitive impairment following moderate to severe TBI.

Overall, failure rates increased as expected with more liberal cut scores. Moreover, findings were consistent with the large body of research showing that RDS < 6 demonstrates excellent specificity (i.e.,>95%) in individuals with moderate to severe TBI undergoing traditional, face-to-face neuropsychological evaluation (see meta-analysis by

Schroeder et al., 2012). Our findings support a cut score of < 7, adding to the conflicting literature about RDS < 7 maintaining adequate specificity (e.g., Larrabee 2003; Etherton et al., 2005; Heinly et al., 2005) or leading to unacceptable false-positive rates (Greiffenstein et al., 1994; Greiffenstein et al., 1995; Babikian et al., 2006; Miller et al., 2012) in moderate to severe TBI. This discrepancy in the literature may be driven by a host of differences across samples that may impact RDS performance including: age, education, current/premorbid intellectual functioning, age at time of injury, time since injury, and presence of potential co-morbid medical or psychiatric conditions. The low PVT failure rate observed by the current sample is consistent with its relatively young age ($M = 38$) and level of education (> 50% beyond high school). Regardless, findings clearly demonstrate that RDS remains robust to the effects of genuine neurocognitive impairment in the context of remote telephone administration.

As with RDS, failure rates rose with increasingly liberal cut-scores across the 21-item test. Findings are consistent with prior research showing that the recommended cut-score of < 8 is associated with near-perfect classification of individuals with objective memory impairments (Iverson et al., 1991; Iverson, Franzen & McCracken, 1994) or neurologic insult (Gontkovsky & Souheaver, 2000) as well as mixed clinical samples (Ryan et al., 2012). In contrast to RDS, research into the use of the 21-item test in the context of moderate to severe TBI is sparse. To date, only one study has investigated the use of the 21-item test in individuals with moderate to severe TBI in the context of the traditional face-to-face testing environment (Rose, Hall, Szalda-Petree & Bach, 1998). Overall performance of the current sample was strikingly similar to the result obtained by Rose and colleagues (1998; $M = 15.4$, $SD = 2.8$). In sum, the telephone administration of the 21-item test does not lead to elevated false positive rates in the context of moderate to severe TBI across the more conservative cut scores.

4.1. Future directions and study limitations

Selection bias may have influenced the results of our study. We relied on a sample of veterans who may be more motivated to perform well when compared to the general clinical population, as evidenced by the sample's willingness to participate in the TBI Model Systems study and to voluntarily complete additional TeleNP and PVT measures. Another limitation is the lack of criterion PVTs beyond those investigated in the current study. As such, it is possible that some of the performances below PVT cutoffs observed in the current sample were indeed invalid (as opposed to false positive classification errors). However, this risk is thought to be minimal given the voluntary nature of the study and the lack of external incentives for poor performance. The veterans in our study were informed their responses to exams would be deidentified, kept confidential for research purposes and, therefore, not be part of the clinical record that could impact future compensation or veterans' benefit determinations. Moreover, none of the veterans in the sample performed below reported cut-scores on both PVTs (i.e., none met criteria for definite malingering). Lastly, the current study utilized TeleNP via telephone. As such, it is unclear the extent to which findings may generalize to other telehealth modalities, such as video-based evaluations.

5. Conclusion

Overall, this study adds to the limited body of research concerning the use of PVTs in TeleNP. To our knowledge, this is the first study to demonstrate that two well-validated PVTs, the RDS and 21-item test, remain robust to the effects of genuine neurocognitive impairment in the context of telephone administration. Moreover, performance on these PVTs by this sample of individuals with moderate to severe TBI was found to be largely consistent with results from studies utilizing TBI samples in traditional face-to-face evaluations. As such, findings provide initial support for the integration of the RDS and 21-item test in TeleNP via telephone. However, future investigations are needed to cross-validate the phone administration of the 21-item test and RDS with other well-established PVTs. Additionally, extensions utilizing either analogue or known-groups design are necessary to fully assess the classification accuracies of the RDS and 21-item test via TeleNP. Specifically, the inclusion of experimental simulators (i.e., adults instructed to feign impairments) or clinical malingerers are necessary to assess the extent to which TeleNP may affect PVT sensitivity. Employing these samples in a counter-balanced cross-over design would allow for experimental investigation of the effects of administration mode (i.e., face to face vs. telephone vs. video-conference) on PVT classification accuracy.

Acknowledgments

We thank Amanda Royer (TBICOE), Jordan Moberg (TBICOE), and Erin Brennan (University of South Florida) and Margaret Wells (Audie L. Murphy Memorial Veterans' Hospital) for assistance with data collection and data management.

Conflict of interest

This research is the result of work supported with resources and the use of facilities at the James A. Haley Veterans' Hospital and Audie L. Murphy Memorial Veterans' Hospital, and was prepared with VHA Central Office VA TBI Model Systems Program of Research and, therefore, is defined as U.S. Government work under Title 17 U.S.C.§101. Per Title 17 U.S.C.§105, copyright protection is not available for any work of the U.S. Government. The authors report no conflicts of interest. The views expressed in this manuscript are those of the authors and do not necessarily represent the official policy or position of the Defense Health Agency, Department of Defense, or any other U.S. government agency. This work was prepared under Contract HT0014-19-C-0004 with DHA Contracting Office (CO-NCR) HT0014 and, therefore, is defined as U.S. Government work under Title 17 U.S.C.§101. Per Title 17 U.S.C.§105, copyright protection is not available for any work of the U.S. Government. For more information, please contact E-mail: dha.TBICOEinfo@mail.mil.

References

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, *20*(1), 145-159.

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R.,... & Woodhouse, J. (2020). InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. *The Clinical Neuropsychologist*, *34*(7-8), 1314-1334.

Brandt, J., Spencer, M., & Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *1*(2), 111-117.

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27*(2), 174-186.

Bunker, L., Hshieh, T. T., Wong, B., Schmitt, E. M., Travison, T., Yee, J., Palihnich, K., Metzger, E., Fong, T. G., & Inouye, S. K. (2017). The SAGES telephone neuropsychological battery: correlation with in-person measures. *International Journal of Geriatric Psychiatry*, *32*(9), 991-999. doi: 10.1002/gps.4558

Castanho, T. C., Amorim, L., Zihl, J., Palha, J. A., Sousa, N., Santos, N. C. (2014). Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: A review of validated instruments. *Frontiers in Aging Neuroscience*, *6*, 16. doi: 10.3389/fnagi.2014.00016

Castanho, T. C., Portugal-Nunes, C., Moreira, P. S., Amorim, L., Palha, J. A., Sousa, N., & Correia Santos, N. (2016). Applicability of the telephone interview for cognitive status (modified) in a community sample with low education level: association with an extensive neuropsychological battery. *International Journal of Geriatric Psychiatry*, *31*(2), 128-136.

Caze, T., Dorsman, K. A., Carlew, A. R., Diaz, A., & Bailey, K. C. (2020). Can You Hear Me Now? Telephone-Based Teleneuropsychology Improves Utilization Rates in Underserved Populations. *Archives of Clinical Neuropsychology*, *35*(8), 1234-1239.

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., . . . Ord, J. S. (2015). Official position of the American academy of clinical neuropsychology social securityadministration policy on validity testing: Guidance and recommendations for change. *The ClinicalNeuropsychologist*, *29*(6), 723-740. doi:10.1080/13854046.2015.1099738

Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Bailey, K. C., Webber, T. A., Alverson, W. A., Messerly, J., González, D. A., & O'Rourke, J. J. F. (2019). Cognitive impairment does not cause performance validity failure: Analyzing performance patterns among unimpaired, impaired, and noncredible participants across six tests. *The Clinical Neuropsychologist*, *33*(6), 1083-1101.

Dillahunt-Aspillaga, C., Nakase-Richardson, R., Hart, T., Powell-Cope, G., Dreer, L. E., Eapen, B. C., Barnett, S., Mellick, D., Haskin, A., & Silva, M. A. (2017). Predictors of employment outcomes in veterans with traumatic brain injury: a VA traumatic brain injury model systems study. *Journal of Head Trauma Rehabilitation*, *32*(4), 271-282.

Donders, J. (2020). The incremental value of neuropsychological assessment: A critical review. *The Clinical Neuropsychologist*, *34*(1), 56-87.

Egeto, P., Badovinac, S. D., Hutchison, M. G., Ornstein, T. J., & Schweizer, T. A. (2019). A systematic review and meta-analysis on the association between driving ability and neuropsychological test performances after moderate to severe traumatic brain injury. *Journal of the International Neuropsychological Society*, *25*(8), 868-877.

Glen, T., Hostetter, G., Roebuck-Spencer, T. M., Garmoe, W. S., Scott, J. G., Hilsabeck, R. C.,... & Espe-Pfeifer, P. (2020). Return on Investment and Value Research in Neuropsychology: A Call to Arms. *Archives of Clinical Neuropsychology*.

Gontkovsky, S. T., & Souheaver, G. T. (2000). Are brain-damaged patients inappropriately labeled as malingering using the 21-Item Test and the WMS-R Logical Memory Forced Choice Recognition Test? *Psychological Reports*, *87*(2), 512-514.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M., 3rd. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Inj*, *15*(12), 1045-1060. doi:10.1080/02699050110088254

Green, R. E., Colella, B., Hebert, D. A., Bayley, M., Kang, H. S., Till, C., & Monette, G. (2008). Prediction of return to productivity after severe traumatic brain injury: investigations of optimal neuropsychological tests and timing of assessment. *Archives of Physical Medicine & Rehabilitation*, 89(12), S51-S60.

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(7), 1093-1129. doi:10.1080/13854040903155063

Hirko, K. A., Kerver, J. M., Ford, S., Szafranski, C., Beckett, J., Kitchen, C., & Wendling, A. L. (2020). Telehealth in response to the COVID-19 pandemic: Implications for rural health disparities. *Journal of the American Medical Informatics Association*, 27(11), 1816-1818.

Iverson, GL. *21-item test manual*. Vancouver, British Columbia, Canada; 1998.

Lacritz, L. H., Carlew, A. R., Livingstone, J., Bailey, K. C., Parker, A., & Diaz, A. (2020). Patient satisfaction with telephone neuropsychological assessment. *Archives of Clinical Neuropsychology*, 35(8), 1240-1248.

Lamberty, G. J., Nakase-Richardson, R., Farrell-Carnahan, L., McGarity, S., Bidelspach, D., Harrison-Felix, C., & Cifu, D. X. (2014). Development of a traumatic brain injury model system within the Department of Veterans Affairs polytrauma system of care. *Journal of Head Trauma Rehabilitation*, 29(3), E1-E7.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32(3), 391-421.

Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical Neuropsychologist*, 1-42.

Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. Archives of Clinical Neuropsychology. Advance online publication. https://doi.org/10.1093/arclin/acaa017

Matchanova, A., Babicz, M. A., Medina, L. D., Rahman, S., Johnson, B., Thompson, J. L., Beltran-Najera, I., Brooks, J., Sullivan, K. L., & Woods, S. P. (2020). Latent Structure of a Brief Clinical Battery of Neuropsychological Tests Administered In-Home Via Telephone. Archives of Clinical Neuropsychology.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol*, 24(8), 1094-1102. doi:10.1076/jcen.24.8.1094.8379

Nakase-Richardson, R., Sherer, M., Seel, R. T., Hart, T., Hanks, R., Arango-Lasprilla, J. C., Yablon, S. A., Sander, A. M., Barnett, S. D., Walker W. C., & Hammond, F. (2011). Utility of post-traumatic amnesia in predicting 1-year productivity following traumatic brain injury: comparison of the Russell and Mississippi PTA classification intervals. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(5), 494-499.

Nelson, L. D., Barber, J. K., Temkin, N. R., Dams-O'Connor, K., Dikmen, S., Giacino, J. T., Kramer, M. T., Levin, H. S., McCrea, M. A., Whyte, J., Bodien, Y. G., Yue, J. K., Manley, G. T., & TRACK-TBI Investigators (2020). Validity of the Brief Test of Adult Cognition by Telephone in Level 1 Trauma Center Patients Six Months Post-Traumatic Brain Injury: A TRACK-TBI Study. *Journal of Neurotrauma*. Advance online publication. doi: 10.1089/neu.2020.7295

Pendlebury, S. T., Welch, S. J., Cuthbertson, F. C., Mariz, J., Mehta, Z., & Rothwell, P. M. (2013). Telephone assessment of cognition after transient ischemic attack and stroke: modified telephone interview of cognitive status and telephone Montreal Cognitive Assessment versus face-to-face Montreal Cognitive Assessment and neuropsychological battery. *Stroke*, 44(1), 227-229. doi: 10.1161/STROKEAHA.112.673384

Plassman, B. L., Newman, T. T., Welsh, K. A., & Helms, M. (1994). Properties of the telephone interview for cognitive status: Application in epidemiological and longitudinal studies. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 7(3), 235-241.

Ramsetty, A., & Adams, C. (2020). Impact of the digital divide in the age of COVID-19. *Journal of the American Medical Informatics Association*, 27(7), 1147-1148.

Ropacki, S., Nakase-Richardson, R., Farrell-Carnahan, L., Lamberty, G. J., & Tang, X. (2018). Descriptive findings of the VA polytrauma rehabilitation centers TBI model systems national database. *Archives of Physical Medicine & Rehabilitation*, 99(5), 952-959.

Rosado, D. L., Buehler, S., Botbol-Berman, E., Feigon, M., León, A., Luu, H.,... & Pliskin, N. H. (2018). Neuropsychological feedback services improve quality of life and social adjustment. *The Clinical Neuropsychologist*, 32(3), 422-435.

Sherer, M., Novack, T. A., Sander, A. M., Struchen, M. A., Alderson, A., & Thompson, R. N. (2002). Neuropsychological assessment and employment outcome after traumatic brain injury: a review. *The Clinical Neuropsychologist*, 16(2), 157-178.

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

Teasdale, G., Maas, A., Lecky, F., Manley, G., Stocchetti, N., & Murray, G. (2014). The Glasgow Coma Scale at 40 years: standing the test of time. *Lancet Neurology*, 13(8), 844-854.

Van Dijk, J. A. (2017). Digital divide: Impact of access. *International Encyclopedia of Media Effects* (pp. 1-11). Medford, MA: John Wiley & Sons. doi: 10.1002/9781118783764.wbieme0043

Weschler, D. (1997). *Weschler Adult Intelligence Scale, Third Edition*. San Antonio, TX: The Psychological Corporation.