Published in final edited form as:

Rehabil Psychol. 2017 November ; 62(4): 455–463. doi:10.1037/rep0000191.

Test Accommodations for Individuals with Neurological Conditions Completing the NIH Toolbox – Cognition Battery: An Evaluation of Frequency and Appropriateness

Susan Magasi,

Departments of Occupational Therapy and Disability Studies, University of Illinois at Chicago

Mark Harniss,

Department of Rehabilitation Medicine, University of Washington

David S. Tulskey,

Center on Assessment Research and Translation and Departments of Physical Therapy and Psychological and Brain Sciences, University of Delaware

Matthew L. Cohen,

Departments of Communication Sciences and Disorders and Psychological and Brain Sciences and Center on Assessment Research and Translation, University of Delaware

Robert K. Heaton, and

Department of Psychiatry, University of California, San Diego School of Medicine

Allen W. Heinemann

Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago

Abstract

Objectives—First, to evaluate the frequency with which individuals with neurological conditions require test administration accommodations for the NIH Toolbox – Cognition Battery (NIHTB-CB). Second, to evaluate the appropriateness of accommodations provided by administrators, including adherence to NIHTB-CB Reasonable Accommodations Guidelines.

Method—Adults with traumatic brain injury, spinal cord injury, or stroke (n=604) completed the NIHTB-CB and other assessments as part of a multi-site study. We provide a descriptive, secondary analysis of test administrator notes to determine use and appropriateness of accommodations.

Results—Four hundred and fifty participants (75%) completed the NIHTB-CB using standard administration procedures, but 137 (22.6%) encountered accessibility challenges which required accommodations. Participants with motor function impairments were most likely to receive at least one of three kinds of accommodations: (a) non-standard methods of entering responses using standard input devices, (b) use of alternate input devices, or (c) directing the test administrator to

enter a response. Fatigue and/or impulsivity led to non-standard administration by 48 (7.9%) individuals. *Post hoc* audit of test administrator notes revealed that despite careful instructions and supervision, 49 (56.3%) of the accommodated administrations breached standardization and scores could not be interpreted using test norms.

Conclusion—Although the NIHTB-CB was developed for individuals without neurological impairment, most individuals with neurological conditions completed the standardized administration without accommodations. When accommodations were needed, administrators did not adhere to the official Reasonable Accommodations Guidelines in more than half of the cases.

Keywords

Neuropsychological Tests Cognition; Spinal cord injuries; Brain injuries; Stroke

Principles of fairness in testing dictate that all test takers have the opportunity to demonstrate their standing on the core constructs that the test is designed to measure (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 2014). Yet, design features and administration protocols can disadvantage test takers with disabilities. Neuropsychological tests typically target cognitive processes and may presume intact, or at least adequate, functioning of the test-takers' visual, auditory, somatosensory, and motor systems, or some combination of these systems. Nearly every cognitive test has motor, vision, and hearing demands that can influence the test score but are not part of the underlying domain being measured (Cicerone et al., 2005; Cicerone et al., 2011; Kettler, 2012). The mode of stimulus presentation and response entry can create barriers for people with motor and sensory impairments (Hill-Briggs, Dial, Morere, & Joyce, 2007). For example, a cognitive test that requires a motor response may create additional demands for people with physical disabilities (Nester, 1994) and a test that involves auditory-verbal instructions or responses may disadvantage people who are deaf or hard of hearing (Braden, 1994). When there is a mismatch between secondary task demands and people's sensorimotor abilities, construct-irrelevant variance may be introduced and artificially lower scores (Messick, 1995, 1998). The Standards for Educational and Psychological Testing identify accessibility of standardized tests as central to issues of fairness in testing and a fundamental validity issue (AERA, APA, & NCME, 2014). Hill-Briggs and colleagues (2007) provide an analysis of how physical and sensory impairments can adversely affect performance on neuropsychological assessments.

Ultimately, it is the responsibility of the assessment professional to select tests that are minimally influenced by examinees' functional limitations. Unfortunately, there are few stand-alone instruments, much less comprehensive batteries that are designed for use with individuals with disabilities.

Adaptations to the original test or the administration process can help increase access for some individuals. Adaptations are typically classified as accommodations or modifications. Accommodations are defined by the Civil Rights Division of the United States Department of Justice as "changes to the regular testing environment and auxiliary aids and services that allow individuals with disabilities to demonstrate their true aptitude or achievement level on

standardized exams or other high-stakes tests” (U.S. Department of Justice, 2014). The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) clarifies that accommodations involve changes in test administration but do not change the underlying construct measured by the instrument, thereby retaining the comparability of scores. In contrast, test *modifications* may alter the construct being measured (AERA, APA, & NCME, 2014; Braden, 2003; Kettler, 2012). The aim of testing accommodations is to shrink the performance gap between people with and without functional impairments. The increase in scores between standard and accommodated versions is called “differential boost” and if accommodations are targeted appropriately, should be greater for people who need the accommodation than those who do not (Fuchs, Fuchs, & Capizzi, 2005; Fuchs et al., 2000). The Americans with Disabilities Act (ADA) ensures that people with disabilities have a right to reasonable accommodations on high stakes testing as long as the accommodations do not result in a fundamental alteration of the test or pose an undue burden on test administrators (U.S. Department of Justice, 2014). However, deviations from the procedure used to create the normative sample could change a test’s psychometric properties (Warschausky et al., 2012) and limit interpretations that reference those normative data.

Cognitive test accommodation and modification remains under studied (Lai & Berkeley, 2012; Sherman & Robinson, 1982). Policies and practices that govern testing accommodations are inconsistent and offer limited guidance to test administrators (Lai & Berkeley, 2012; Mislevy et al., 2013). Furthermore, decisions about accommodations often need to be made “in the moment” based on the unique needs of an individual, and test administrators may have limited training and understanding of disability (American Psychological Association [APA], 2012).

The NIH Toolbox for the Assessment of Neurological and Behavioral Function (NIHTB) is a comprehensive set of standardized, performance-based measures of cognitive, motor, and sensory function and emotional health (Gershon et al., 2010). It capitalizes on item-response theory and computerized adaptive testing and does not require extensive administrator training. Because of its sophistication and support from NIH, it sets a standard for clinical and epidemiological research (Quatrano & Cruz, 2011). The current investigation focuses on the NIHTB - Cognition Battery (NIHTB-CB), which is comprised of 7-computer administered measures of cognition, and is the product of rigorous development, validation and national standardization (Weintraub et al., 2014).

Through its development, the NIHTB team integrated principles of accessibility and universal design to enhance NIHTB’s accessibility to the general population aged 3 to 85 years. The NIHTB development team included an Accessibility Work Group that reviewed all measures for accessibility and usability by people with physical, cognitive, and sensory impairments and provided recommendations to enhance accessibility (Magasi, Harniss, & Heinemann, 2017; Rios, Magasi, Novak, & Harniss, 2016; Victorson et al., 2013). Figure 1 summarizes accessibility of measures for specific impairment groups.

The Accessibility Work Group developed Reasonable Accommodation Guidelines for each test in the NIHTB-CB (NIH Toolbox, 2012). The purpose of this study is (1) to describe the

frequency of accommodations that were required by individuals with three neurological conditions, traumatic brain injury (TBI), spinal cord injury (SCI), or stroke; and (2) to evaluate the appropriateness of the accommodations, that is, how consistently test administrators applied the recommendations in the Reasonable Accommodations Guidelines.

Methods

Design/Setting

Cross-sectional data were collected as part of a study aimed at validating the NIHTB in people with SCI, TBI and stroke (Tulsky & Heinemann, in preparation; Tulsky et al., under review). Assessment occurred at two academic medical centers and a free-standing rehabilitation hospital in the Midwestern United States (St. Louis, Ann Arbor, and Chicago). Collaborating sites obtained ethics approval.

Participants

The sample included 604 people with SCI, TBI, or stroke who were recruited through community and clinically-based networks. Eligibility criteria included community residence; medically confirmed diagnosis of traumatic SCI, stroke, or TBI; at least one-year following their most recent injury; 18 to 85 years of age; ability to read at a fifth-grade reading level; and ability to speak English. Participants provided informed consent, completed the Lighthouse Near Visual Acuity Test to ensure visual acuity of at least 20/100, and the Wide Range Achievement Test, Fourth Edition (WRAT-IV) Reading test to estimate reading level. Participants received an honorarium of US \$90 per day for their time and contribution to the research.

Data Collection

Testing occurred over 2 days lasting an average of 4.9 and 4.7 hours per day, respectively. Administrators followed a fixed NIHTB-CB test order under guidance of a certified test administrator on the first day. Test administrators were trained rigorously and certified to administer the NIHTB-CB per standardized procedures and received supplemental training on the importance and application of the Reasonable Accommodations Guidelines. Test administrators were re-certified after one year to minimize drift in the administration protocol. The project coordinator met bi-weekly with the administrators to ensure consistent procedures across sites.

Administrators were instructed to record for each participant whether they performed a complete and standardized administration of the NIHTB-CB, and if not, to record details of the non-standard administration. Although administrator records were not intended to be used for research purposes, based on in-depth review of the quality and consistency of these records, we had sufficient confidence in the data set to perform qualitative reviews of administrator records for this investigation. All data were collected between December 2011 and November 2013.

Analysis

Two coders with expertise in disability accommodations and cognitive testing (authors SM and MH) independently coded all administrator records using a stepwise process and a common coding dictionary. Coders met regularly to compare their categorizations and reconcile discrepancies to promote a consistent coding structure (Saldaña, 2013). First, the coders sorted administrations based on whether they were *standard* (i.e., adhered to NIHTB-CB administration protocols) or *non-standard*. *Non-standard* administrations were then categorized by the reason for non-standard administration, *accessibility issues* or *technical/logistical issues*. *Accessibility issues* were categorized as: (a) *received accommodations*, (b) *Unable to complete* as defined by NIHTB inclusion/exclusion criteria for a measure (e.g., unable to complete the picture vocabulary measure due to deafness or low vision), and (c) *secondary impairment effects* that may have influenced test performance (e.g., fatigue, distractibility, and impulsivity), a category not identified in the Reasonable Accommodation Guidelines.

Coders categorized *technical/logistical issues* as *computer issues* (e.g., technical issues that prevented the computer-based measures from performing as expected) or *study logistical issues* (e.g., participant came late for the appointment or left early).

Finally, the coders reviewed all descriptions of non-standard administrations and compared them with the Reasonable Accommodations Guide to identify deviations. They classified non-standard administrations as: (a) accommodation does not alter core task demand (keep score), (b) accommodation alters core task demands (drop score), (c) unclear impact on core task demands (interpret score with caution), or (d) appropriate omission. A third rater independently verified classifications, a neuropsychologist with expert knowledge of the NIHTB-CB development, administration and scoring (see acknowledgements NC). The third rater agreed with the consensus categorizations in all cases.

Results

Table 2 provides the demographic and clinical characteristics of the sample. According to the administrator records, 450 participants (75%) received a standardized administration of the entire NIHTB-CB and 154 participants (25%) received a non-standard administration on at least one NIHTB-CB measure (Figure 2). Accessibility issues accounted for 137 (89%) of non-standard administrations and technical/logistical reasons accounted for the remainder (17, 11%). Technical/logistical issues included computer issues (10, 59%), logistical issues (4, 23%), and other reasons (3, 18%).

Of the 137 individuals who experienced accessibility issues, 87 (64%) required accommodations, 48 (35%) displayed secondary impairment effects that may have influenced performance but were not addressed by the Reasonable Accommodations Guidelines, and 2 (<1%) were unable to complete the tasks and discontinued the battery. Accommodated administrations were provided most commonly to people with motor impairments, including hemiplegia and hemiparesis secondary to stroke, motor paralysis secondary to SCI, and apraxia secondary to TBI. Modified administration protocols can be classified into three categories: (a) non-standard ways of entering responses using standard

input devices, e.g. depressing arrow keys with non-dominant hand or use of alternate motor patterns, such as a knuckle, side of the hand or middle finger in lieu of isolated movement of the index finger on dominant hand, (b) use of alternate input devices, such as touchscreen or track ball; and (c) participant directing test administrator to enter response. Administrators provided verbal cuing when participants needed reminders of aspects of the administration protocol but were otherwise able to perform the task independently.

Secondary impairment issues fell into two categories. The first category included sensory impairments, such as decreased vision or hearing, which limited the test taker's ability to perceive test stimuli or response options. In these instances, administrators appropriately omitted tests from the administration battery. The second category included secondary impairment effects, such as impulsivity and fatigue, which limited the test taker's ability to attend to the testing process. This category of non-standard administrations is not described in the Reasonable Accommodations Guidelines, but emerged as a significant factor that influenced participants' abilities to comply with the standardized administration protocols.

The "unable to complete" category refers to omission of the entire cognition battery, rather than omission of individual measures. Exclusions were the result of participants' decisions to terminate testing because of frustration or failure.

After classifying non-standard administrations, the coders reviewed the administrator records for the 87 accommodated administrations and compared their practices with the Reasonable Accommodations Guidelines. The coders paid particular attention to the core and secondary task demands of each NIHTB-CB test, as well as the components of each test's scoring algorithm. For example, a response entry by the test administrator acting under the direction of a participant may be acceptable for the NIHTB-CB Picture-Sequence Memory test in which effortful motor control required for "dragging and dropping" small pictures around the screen is irrelevant to the core construct of the test (episodic memory) and does not influence the scoring algorithm (i.e., scoring is not based on speed). In contrast, it would be a significant breach of standardization for the test administrator to enter participant's responses on a timed test, such as the NIHTB-Pattern Comparison Test. Of the 87 non-standard administrations that were classified as receiving accommodations, expert review indicated that 49 (56.3% or 8.1% of the total sample) breached standardization thereby invalidating test scores. Additionally, 21 of the 137 non-standard administrations (15.3%) were identified for potential impacts on interpretation because the accommodation may have compromised normative interpretation of the scores. Without empirical evaluation, it is not possible to determine the consequence of the accommodations. We recommended that these scores be interpreted with caution. The most common reasons for this designation was when fatigue or impulsivity compromised performance. Table 3 shows the frequency of appropriate and inappropriate accommodations.

Discussion

This study evaluated the frequency and appropriateness of test accommodations provided to people with SCI, TBI and stroke within the context of a validation study of the NIHTB-CB (Tulsky & Heinemann, in process). It contributes to our understanding of how test

administrators make accommodations for people with a range of physical, cognitive, and sensory impairments on cognitive tests. With appropriate training, certification, guidance and supervision, test administrators provided standardized administrations to most (75%) of the sample. Individuals who required accommodations most often needed alternate modes of response entry. Fatigue and impulsivity emerged as issues that require accommodation.

We also examined the fidelity of examiners in tailoring accommodations to the Reasonable Accommodations Guidelines. Inappropriate accommodations that precluded comparison with the normative sample occurred in 56.3% of the accommodated administrations. Although we considered the possibility that nonadherence reflects inadequate training or supervision of examiners, we dedicated far more attention to this issue than is typical. Instead, we believe that nonadherence reflects the limited experience that most test administrators have in ensuring that all participants had an “unobstructed opportunity to demonstrate their standing on the construct being measured.” Adherence to APA Guidelines for Assessment of and Intervention with Persons with Disabilities (APA, 2012) requires content knowledge, experience, and judgement to inform decisions, which some research assistants and test administrators lack. Most test instructions do not include procedures for reasonable accommodations. Although the NIHTB-CB does have a Reasonable Accommodations Guide, it is not integrated into the computer administration platform. As a consequence, administrators make “in the moment” decisions without access to the Guide.

Results highlight the need for careful and on-going scrutiny of test accommodations and empirical evaluation of the consequences that accommodations have on cognitive test scores. Evaluation of reasonable accommodations on test scores has been challenging due to the heterogeneity of the disability population creates challenges in disability accommodations (Lai & Berkeley, 2012). Although research on accommodation effectiveness has grown steadily since the passage of the ADA, it remains relatively sparse and has been conducted primarily within the context of educational and high-stakes testing (Kettler, 2012; Kettler, Elliott, & Beddow, 2009). Additional research into how accommodations can minimize barriers and reduce construct irrelevant variance is needed (Haladyna & Downing, 2004; Lai & Berkeley, 2012).

Findings also underscore that secondary conditions, such as fatigue and impulsivity, may reduce the ability of some individuals to comply with standardized administration protocols. We identified 21 participants whose test performance may have been influenced by secondary conditions; however, this may be a low estimate. Because this category of nonstandard administrations was not articulated in the NIHTB Reasonable Accommodations Guide, not all test administrators commented on these issues or provided detailed documentation that could be scored by the raters. It is possible that secondary impairments occurred at a greater rate than reported here. While there is a body of literature that demonstrates that fatigue is not related to performance on cognitive and neuropsychological assessments (Ashman et al., 2008; Zhu & Tulsy, 2000), these findings suggest that fatigue makes it harder for people to comply with standardized administration protocols. Fatigued individuals may require verbal cuing to stay on task. Their behavior, and the administrator’s response to their behavior, may compromise the validity of test score interpretation. Test

administrators should record verbal and non-verbal signs of distraction, fatigue, and frustration.

Our findings support the American Psychological Association's and American Educational Research Association's 2012 Guidelines for the Assessment of and Intervention with Persons with Disabilities, which seek to optimize fairness in testing while maintaining validity. While some accommodations may breach standardization, "for many kinds of tests, an accommodated measure would yield more valid results than the same measure without such accommodations" (APA, 2012, p. 53). When tests and administration protocols are modified, new normative information must be collected.

It may be possible in some cases to substitute comparable measures that do not introduce access barriers for test takers and construct irrelevant variance to test scores (Haladyna & Downing, 2004; Hollenbeck, 2005). For example, Carlozzi and colleagues (under review) developed motor-free versions of the NIHTB-CB for people with upper extremity impairments. While the hands-free version is not a perfect match with the original, they provide a reliable and valid approach to measuring global and fluid cognition for people who would not otherwise be able to participate.

Limitations

Post hoc analysis of examiner notes revealed accommodation procedures that deviated from the Reasonable Accommodations Guidelines, but we are unable to evaluate the effects that these deviations had on test scores and the validity of accommodated administrations. Although the analyses are only descriptive, they illustrate the consequences of barriers in the NIHTB-CB administration procedures and the application of reasonable accommodations on the completeness and validity of the data. Although we instructed examiners to document test procedures consistently, they did not use a standardized reporting format, which resulted in richer behavioral descriptions from some examiners than others. Future projects should systematize the documentation of test administration deviations.

Future studies should evaluate the effects of accommodations on test scores and the psychometric properties of tests. Well controlled test accommodation studies that compare standard and accommodated administration of parallel tests of people with and without disabilities would enable equivalence testing and direct measurement of differential boost between people with and without disabilities (Hansen, Mislevy, Steinberg, Lee, & Forer, 2005; Hollenbeck, 2005; Kettler, 2012; Kettler et al., 2009). These studies could improve our understanding of how accessibility and accommodations affect summary scores of test scores (Heaton et al., 2014).

Clinical implications

The use of appropriate and well-documented accommodations ensures that principles of fairness in testing are integrated into the test administration process while minimizing breaches in standardization. Planning for access helps ensure a better experience for test administrators and test takers with disabilities, while maintaining data quality. We recommend the following steps to enhance access and accommodations when testing people with disabilities.

Targeted Recommendations for Future Development of the NIHTB-CB

- Consistent integration of the Reasonable Accommodation Guidelines within the NIHTB administration platform would promote ease of access, including enhanced organization and "searchability" to facilitate ease of use within the testing session.
- Alternative administration instructions and procedures for certain tests would allow more individuals to be tested.
- Continuous review of the Reasonable Accommodations Guidelines can help ensure that new challenges are identified and addressed promptly and appropriately.

Targeted Recommendation for Users of the NIHTB-CB

- Test administrators need rigorous training on how to work with people with functional, physical, cognitive and sensory impairments. Training requirements are the same for the NIHTB-CB as for any test of cognition. Administrators should review APA's Guidelines for Assessment of and Intervention with Persons with Disabilities (2012), particularly Guideline 13 (disability as a dimension of diversity), 14 (psychometric soundness, fairness, comprehensiveness, and appropriateness), and 15 (appropriate accommodations).
- Diversity of disability samples may require targeted training and monitoring. While our team scheduled bi-weekly meetings, agendas emphasized data collection of the 2-day assessment battery rather than access and accommodation issues. Purposeful and explicit review of these issues can ensure that breaches in standardization and the use of inappropriate accommodations are caught early and study-wide strategies are implemented to ensure data quality.
- We recertified test administrators annually to minimize drift from standardized administration protocols. More frequent monitoring of administration practices may be required for research and clinical applications.
- Within the context of clinical research, it is important to acknowledge the pressures placed on test administrators and research assistants to obtain complete data sets. Explicit and frequent decisional support is needed to help administrators know when to provide accommodations, when accommodations breach standardization and when measures are appropriately omitted than accommodated.
- A static posture during testing can lead to fatigue and somatic complaints such as pain and spasticity. Test administrators should monitor signs of fatigue, impulsivity, pain and spasticity when working with people with SCI, TBI and stroke. Administrators should provide short rest or reposition breaks to minimize the consequences of secondary impairments on test performance.

General Recommendations for the Development and Use of Computer-Based Assessments with People with Disabilities

- Test developers should develop universally designed assessments to ensure access to the broadest range of people possible (AERA, APA, & NCME, 2014; Meyers & Andresen, 2000).
- Test developers should define and document the core construct the test is designed to measure (AERA, APA, & NCME, 2014).
- Test developers should define acceptable accommodations in advance of testing, including the identification of appropriate alternative measures when available.
- Given that 20% of Americans have a disability (Brucker & Houtenville, 2015), all test administrators are likely to encounter people with disabilities (Rios et al., 2016), yet few test administrators receive training in how to work effectively with this population (APA, 2012). Test administrators should receive training in how to work with people with a range of functional, physical, cognitive and sensory impairments.
- Test administrators should recognize the importance of standardization and application of appropriate accommodations in order to make informed decisions in testing sessions.
- Test administrators should understand the core constructs that tests are designed to measure, as well as the secondary test demands inherent in the administration process.
- *Post hoc* review is required to determine if test accommodations breached standardization and compromised the validity of scores. Team leaders with an understanding of the core constructs and the consequences of accommodations should assume responsibility for a review process.
- In spite of accommodations' uncertain consequences on test scores, "flagging" test scores derived from accommodated administrations is prohibited as a discriminatory practice under the ADA (U.S. Department of Justice, 2014). Validity reflects the context of test use. Different contexts may demand varying levels of methodological rigor and adherence to standardized administration protocols. For example, high stakes testing requires a level of rigor and standardization than may be needed in other contexts. Clinicians and researchers should balance the need for flexibility with adherence to test standardization.

Conclusion

Neuropsychological assessment involves balancing accommodations of participants' abilities and adherence to standardized procedures. Accommodations can improve accessibility, but are most effective and the least confounding when they are constrained by guidelines for a test. The NIHTB-CB Reasonable Accommodations Guide provide a good model. Here we reported on departures from these Guidelines despite rigorous training and

supervision practices, and our findings can inform future revisions of the Guide. Given that many research studies are implemented without predefined testing accommodations and with less extensive training and monitoring, our findings suggest that researchers should pay attention to the decision making of test administrators and the un-documented accommodations they may be using.

Acknowledgments

This research was supported by grant #H133B090024 (PI: Heinemann) from the National Institute on Disability, Independent Living, and Research Rehabilitation (NIDILRR), as well as a Psychosocial Research Postdoctoral Fellowship Grant from the Craig H. Neilsen Foundation (#364962; PI: Cohen). Salary support for Dr. Cohen was also provided by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number U54-GM104941 (PI: Binder-Macleod). The authors would like to thank their talented and dedicated research staff: Ana Miskovic, Arielle Goldsmith, Sara Jerousek, Sue Tucker, Jessica Dashner, Cathy Crawley, Ben Aziz, Julie Grech, Livia Vanden Belt, and Angela Miciura. The authors would like to thank Noelle Carlozzi for her expert review of the coded data and Kimberley The for her editorial support.

References

- AERA, APA, & NCME. STANDARDS for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
- American Psychological Association. Guidelines for assessment of and intervention with persons with disabilities. *American Psychologist*. 2012; 67:43–62. DOI: 10.1037/a0025892 [PubMed: 22103887]
- Ashman TA, Cantor JB, Gordon WA, Spielman L, Egan M, Ginsberg A, Flanagan S. Objective measurement of fatigue following traumatic brain injury. *Journal of Head Trauma Rehabilitation*. 2008; 23:33–40. DOI: 10.1097/01.htr.0000308719.70288.22 [PubMed: 18219233]
- Braden, JP. Deafness, deprivation, and IQ. New York, NY: Plenum Press; 1994.
- Braden, JP. Accommodating clients with disabilities on the WAIS-III and WMS-III. In: Tulskey, DS., et al., editors. *Clinical interpretation of the WAIS-III and WMS-III*. San Diego, CA: Academic Press; 2003. p. 451–486.
- Brucker DL, Houtenville AJ. People with disabilities in the United States. *Archives of Physical Medicine and Rehabilitation*. 2015; 96:771–774. DOI: 10.1016/j.apmr.2015.02.024 [PubMed: 25757791]
- Carlozzi NE, Goodnight S, Umlauf A, Heaton R, Heinemann A, Schalet BD, Tulskey D. Motor-Free composites from the NIH Toolbox Cognition Battery for people with disabilities. *Rehabilitation Psychology*. (under review).
- Cicerone KD, Dahlberg C, Malec JF, Langenbahn DM, Felicetti T, Kneipp S, Catanese J. Evidence-based cognitive rehabilitation: updated review of the literature from 1998 through 2002. *Archives of Physical Medicine and Rehabilitation*. 2005; 86:1681–1692. DOI: 10.1016/j.apmr.2005.03.024 [PubMed: 16084827]
- Cicerone KD, Langenbahn DM, Braden C, Malec JF, Kalmar K, Fraas M, Bergquist T. Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. *Archives of Physical Medicine and Rehabilitation*. 2011; 92:519–530. DOI: 10.1016/j.apmr.2010.11.015 [PubMed: 21440699]
- Fuchs LS, Fuchs D, Capizzi AM. Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children*. 2005; 37(6):1–8.
- Fuchs LS, Fuchs D, Eaton SB, Hamlett C, Binkley E, Crouch R. Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*. 2000; 67(1):67–81.
- Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. *The Lancet Neurology*. 2010; 9:138–139. DOI: 10.1016/S1474-4422(09)70335-7 [PubMed: 20129161]
- Haladyna TM, Downing SM. Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*. 2004; 23:17–27. DOI: 10.1111/j.1745-3992.2004.tb00149.x

- Hansen EG, Mislevy RJ, Steinberg LS, Lee MJ, Forer DC. Accessibility of tests for individuals with disabilities within a validity framework. *System*. 2005; 33:107–133. DOI: 10.1016/j.system.2004.11.002
- Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, Gershon R. Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society*. 2014; 20:588–598. DOI: 10.1017/s1355617714000241 [PubMed: 24960398]
- Heinemann AW, Magasi S, Hammel J, Carlozzi NE, Garcia SF, Hahn EA, Jerousek S. Environmental factors item development for persons with stroke, traumatic brain injury, and spinal cord injury. *Archives of Physical Medicine and Rehabilitation*. 2015; 96:589–595. DOI: 10.1016/j.apmr.2013.11.024 [PubMed: 24378804]
- Hill-Briggs F, Dial JG, Morere DA, Joyce A. Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness. *Archives of Clinical Neuropsychology*. 2007; 22:389–404. DOI: 10.1016/j.anc/2007.01.013 [PubMed: 17303374]
- Hollenbeck K. Validity Issues and decisions about testing accommodations. *Assessment for Effective Intervention*. 2005; 31:7–17. DOI: 10.1177/073724770503100102
- Kettler RJ. Testing Accommodations: Theory and research to inform practice. *International Journal of Disability, Development & Education*. 2012; 59:53–66. DOI: 10.1080/1034912X.2012.654952
- Kettler RJ, Elliott SN, Beddow PA. Modifying achievement test items: A theory-guided and data-based approach for better measurement of what students with disabilities know. *Peabody Journal of Education*. 2009; 84:529–551. DOI: 10.1080/01619560903240996
- Lai SA, Berkeley S. High-stakes test accommodations: Research and practice. *Learning Disability Quarterly*. 2012; 35:158–169. DOI: 10.1177/0731948711433874
- Magasi S, Harniss M, Heinemann AW. An interdisciplinary approach to the development of accessible computer-administered measurement instruments. *Archives of Physical Medicine & Rehabilitation*. (in press).
- Messick SA. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50:741–749. DOI: 10.1037/0003-066x.50.9.741
- Messick SA. Test Validity: A Matter of Consequence. *Social Indicators Research*. 1998; 45:35–44. DOI: 10.1023/A:1006964925094
- Meyers AR, Andresen EM. Enabling our instruments: Accommodations, universal design, and access to participation in research. *Archives of Physical Medicine and Rehabilitation*. 2000; 81:55–59. DOI: 10.1053/apmr.2000.20618
- Mislevy RJ, Haertel G, Cheng BH, Ructtinger L, DeBarger A, Murray E, Rutstein D. A “conditional” sense of fairness in assessment. *Educational Research and Evaluation*. 2013; 19:121–140. DOI: 10.1080/13803611/2013.767614
- NIH Toolbox. Reasonable Accommodations Guidelines. 2012. Retrieved from http://www.healthmeasures.net/images/nihtoolbox/Accessibility_Manuals/NIH_Toolbox_Reasonable_Accommodations_Guidelines.pdf
- Nester, MA. Psychometric testing and reasonable accommodations for persons with disabilities. In: Bruyere, SM., O'Keefe, J., editors. *Implications of the Americans with Disabilities Act for Psychology*. New York, NY: Springer Publishing Company and American Psychology Association; 1994. p. 25–36.
- Quatrano LA, Cruz TH. Future of outcomes measurement: Impact on research in medical rehabilitation and neurologic populations. *Archives of Physical Medicine and Rehabilitation*. 2011; 92:S7–S11. DOI: 10.1016/j.apmr.2010.08.032 [PubMed: 21958925]
- Rios D, Magasi S, Novak C, Harniss M. Conducting accessible research: Including people with disabilities in public health, epidemiological, and outcomes studies. *American Journal of Public Health*. 2016; 106:2137–2144. DOI: 10.2105/AJPH.2016.303448 [PubMed: 27736212]
- Saldaña, J. *The Coding Manual for Qualitative Researchers*. 2. Los Angeles, CA: Sage;
- Sherman, SW., Robinson, NM. *Ability Testing of Handicapped People: Dilemma for Government, Science, and the Public*. Washington, DC: National Academy Press; 1982.

- Tulsky DS, Heinemann AW. Validating the NIH Toolbox - Cognitive Battery in individuals with Disabilities. *Rehabilitation Psychology*. (In Preparation).
- Tulsky DS, Carlozzi NE, Holdnack JA, Heaton RK, Wong A, Goldsmith A, Heinemann AW. Using the NIH Toolbox Cognition Battery (NIHTB-CB) in individuals with a traumatic brain injury. *Rehabilitation Psychology*. (Under Review).
- U.S. Department of Justice. ADA requirements - Testing accommodations. 2014. Retrieved from https://www.ada.gov/regs2014/testing_accommodations.pdf
- Victorson D, Manly J, Wallner-Allen K, Fox N, Purnell C, Hendrie H, Gershon R. Using the NIH Toolbox in special populations: Considerations for assessment of pediatric, geriatric, culturally diverse, non-English-speaking, and disabled individuals. *Neurology*. 2013; 80:S13–19. DOI: 10.1212/WNL.0b013e3182872e26
- Warschawsky S, Van Tubbergen M, Asbell S, Kaufman J, Ayyangar R, Donders J. Modified test administration using assistive technology: Preliminary psychometric findings. *Assessment*. 2012; 19:472–479. DOI: 10.1177/1073191111402458 [PubMed: 21467093]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, Gershon R. The cognition battery of the NIH Toolbox for Assessment of Neurological and Behavioral Function: Validation in an adult sample. *Journal of the International Neuropsychological Society*. 2014; 20:567–578. DOI: 10.1017/S1355617714000320 [PubMed: 24959840]
- Zhu J, Tulsky DS. Co-norming the WAIS-III and WMS-III: Is there a test-order effect on IQ and memory scores? *Clinical Neuropsychologist*. 2000; 14:461–467. DOI: 10.1076/clin.14.4.461.7197 [PubMed: 11262715]

Impact

- This is the first study to report on the frequency and appropriateness of accommodations provided to people with neurological conditions who are completing the NIH Toolbox – Cognitive Battery.
- While the majority (75%) of people with neurological conditions could complete the NIHTB-CB using standard administration procedures, 22.6% of the sample and 56.3% of the accommodations breached standardization to the extent that validity of test scores were threatened.
- Consistent application of reasonable accommodations can minimize data loss due to breached standardization and application of inappropriate accommodations. Case-by-case, *post hoc* analysis by accessibility and content experts can help ensure the integrity of data used in clinical and research applications.

	Vision		Physical and Motor Skills		Hearing	Language & Learning Aa		
Measure	Blind	Low Vision	Decreased Hand Function	Mobility Impairment	Deaf	Hard of Hearing	Speech impairment	Reading Impairment
NIH Toolbox Picture Vocabulary Test	N	P	P	A	P	P	A	A
NIH Toolbox Flanker Inhibitory Control and Attention	N	P	P	A	P	A	A	A
NIH Toolbox List Sorting Working Memory Test	N	P	A	A	A	A	N	A
NIH Toolbox Dimensional Change Card Sort	N	P	P	A	A	A	A	P
NIH Toolbox Pattern Comparison Processing Speed	N	P	P	A	P	A	A	A
NIH Toolbox Picture Sequence Memory Test	N	P	P	A	P	A	A	A
NIH Toolbox Oral Reading Recognition Test	N	P	A	A	P	A	N	A

N= not accessible and not feasible to make accessible
P=Not accessible, but can be made accessible with reasonable accommodations
A= Currently accessible

Figure 1.
Accessibility of NIHTB Cognition Battery for specific impairment groups

Figure 2.
Categorized administration of the NIHTB Cognition Battery in a cohort of 604 people with SCI, TBI and Stroke.

Table 1**NIHTB Cognition Battery – Description of Core Constructs and Task Demands**

NIH Toolbox Test	Core Construct	Administration Protocol
Picture Vocabulary Test	Language: Receptive Vocabulary	<p>Instructions: Onscreen text and verbal by administrator</p> <p>Presentation: Audio Cue Words; Visual Images of response options</p> <p>Response Format: Mouse click on visual image</p> <p>Timing: Self-paced</p>
Flanker Inhibitory Control & Attention Test	Executive Function: Inhibitory Control and Attention	<p>Instructions: On screen text and verbal by administrator</p> <p>Presentation: Visual Images</p> <p>Response Format: Keyboard arrow keys w/ index finger of dominant hand</p> <p>Timing: Response time included in scoring algorithm.</p>
Dimensional Change Card Sort Test	Executive Function: Set-Shifting	<p>Instructions: On screen text and verbal by administrator</p> <p>Presentation: Visual presentation of graphical instrument content</p> <p>Onscreen text of cue word</p> <p>Response Format: Keyboard arrow keys w/ index finger of dominant hand</p> <p>Timing: Response time included in scoring algorithm</p>
Pattern Comparison Processing Speed Test	Processing Speed	<p>Instructions: Verbal by administrator</p> <p>Presentation: Visual presentation of graphical images.</p> <p>Response Format: Keyboard arrow keys w/ index finger of dominant hand</p> <p>Timing: Response time included in scoring algorithm</p>
List Sorting Working Memory Test	Working Memory	<p>Instructions: Verbal by administrator</p> <p>Presentation: Visual images, on screen text and audiorecorded</p> <p>Response Format: Verbal by participant.</p> <p>Timing: Self-paced</p>
Picture Sequencing Memory Test	Episodic Memory	<p>Instructions: Verbal by administrator</p> <p>Presentation: Visual images and audio description</p> <p>Response Format: Mouse click on visual image to select and place</p> <p>Timing: Self-paced</p>
Oral Reading Recognition Test	Language: Oral Reading	<p>Instructions: Verbal by administrator</p> <p>Presentation: Written Text</p>

NIH Toolbox Test	Core Construct	Administration Protocol
		Response Format: Verbal by participant Timing: Self-paced

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Participant Demographic and Clinical Information

	SCI	TBI	Stroke	Total
	N=209	N=184	N=211	N=604
Age (in years), mean (SD)	46(14)	40(17)	56(130)	47(16)
Gender				
Male	164 (78%)	118 (64%)	105 (50%)	387 (64%)
Female	45 (22%)	66 (36%)	106 (50%)	217 (36%)
Ethnicity				
Hispanic	18 (9%)	13 (7%)	11 (5%)	42 (7%)
Race				
White	128 (61%)	135 (74%)	91 (43%)	354 (59%)
Black or African American	60 (29%)	29 (16%)	104(50%)	193 (32%)
Asian	1 (1%)	9 (5%)	4 (2%)	14 (2%)
Other	20 (10%)	11 (7%)	12 (6%)	43 (23%)

Table 3

Non-Standard Administrations by Test

NIH Toolbox Test	Appropriate Omission	Inappropriate Accommodation	Questionable Accommodation	Total
Picture Vocabulary	1	3	10	14
Flanker Inhibitory Control & Attention	6	39	16	61
Dimensional Change Card Sort	6	41	15	62
Pattern Comparison	6	38	13	57
List Sorting Working Memory	1	2	7	10
Picture Sequence Memory	0	5	11	16
Reading Recognition	1	2	8	11

Appropriate Omission - based on review of administrator notes, test appropriately omitted during TB administration

Inappropriate Accommodation - based on review of administrator notes, these administrations represent significant breaches in standardization, most commonly use of alternate input devices touchscreen, mouse, mouth stick, or research assistant entry of responses

Questionable Accommodation - based on review of administrator notes, these administration do not alter core concept but accommodations may limit interpretability of derived scores, usually had to do with use motor patterns other than isolated movement of dominant index finger or impact of secondary conditions such as fatigue and impulsivity