Published in final edited form as:

Appl Neuropsychol Adult. 2018 ; 25(6): 489–496. doi:10.1080/23279095.2017.1329147.

Comparing the Test of Practical Judgment with the Neuropsychological Assessment Battery Judgment Subtest in a Neurodegenerative Disease Clinic Population

Ms January Durant,

Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W Bonneville Ave, Las Vegas, 89106
United States

Ms Jody-Lynn Berg,

Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W Bonneville Ave, Las Vegas, 89106
United States, bergj@ccf.org

Dr Sarah Jane Banks, and

Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W Bonneville Ave, Las Vegas, 89106
United States, bankss2@ccf.org

Dr Justin Benjamin Miller

Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W Bonneville Ave, Las Vegas, 89106
United States, millerj4@ccf.org

Abstract

Objective: The Test of Practical Judgment (TOP-J) and the Judgment subtest from the Neuropsychological Assessment Battery (NAB-JDG) are both brief interview based measures that assess judgment. This study compared estimates of judgment obtained from these measures in a neurodegenerative disease population.

Method: Records from 61 referrals seen for neuropsychological evaluation in a neurodegenerative disorders clinic were reviewed. Measures of interest included the TOP-J and NAB-JDG. Concordance correlation coefficients (CCC) and root mean square differences (RMSD) were calculated between judgment T-scores. Discrepancy scores were calculated by subtracting NAB-JDG scores from TOP-J scores.

Results: CCC showed poor agreement between the judgment measures, with evidence of fixed bias, such that the NAB-JDG systematically generates higher scores than the TOP-J. This fixed bias was present whether NAB-JDG scores are demographically adjusted or unadjusted. There was no evidence of proportional bias.

Conclusions: In a neurodegenerative disease clinic population, the TOP-J and NAB-JDG provide estimates of judgment ability that are systematically different. These two measures may be assessing different aspects within the larger construct of judgment and the inconsistency between measures would contraindicate using them interchangeably. Clinicians will need to carefully

consider patient characteristics, clinical needs, and review specific item content when selecting between these measures.

Keywords

judgment; neurodegenerative disease; Test of Practical Judgment; Neuropsychological Assessment Battery, concordance correlation coefficient

Introduction

Impairments in judgment and decision-making ability are associated with the progressive cognitive decline that occurs in neurodegenerative diseases such as Alzheimer's disease (AD) and other types of dementia (The Alzheimer's Association, 2015). Judgment may be defined as the ability to carefully evaluate a situation and make an appropriate decision based on relevant information, context, potential solutions and outcomes (Capucho & Brucki, 2011; Rabin et al., 2007). Judgment is a complex cognitive skill that is often conceptualized within the domain of executive functioning (Strauss, Sherman, & Spreen, 2006). Impairments in the ability to make sound decisions may pose safety concerns, impact an individual's ability to function independently, and indicate the need for increased supports. Given the importance of judgment ability in daily living, formal assessment of this domain is often an important component of a neuropsychological assessment battery, particularly for individuals suspected of neurodegenerative disorders.

Rabin, Borgos, and Saykin (2008) conducted a survey examining neuropsychologists' practices and perspectives regarding the assessment of judgment ability. Results indicated 87% of the neuropsychologists surveyed often or always assess judgment when evaluating patients with dementia and 82% often or always use formal psychological tests. The neuropsychologists reported using a variety of tests to evaluate judgment ability, with the most commonly used instruments being tests that were not specifically designed to assess judgment such as the Wechsler Adult Intelligence Scale, Third Edition Comprehension subtest (WAIS-III; Wechsler, 1997), Wisconsin Card Sorting Test (WCST; Heaton et al., 1993), and WAIS-III Similarities subtest (Wechsler, 1997).

Despite the importance of assessing judgment ability in older adults, there are a relatively limited number of neuropsychological tests designed explicitly for this purpose (Capucho & Brucki, 2011). Some judgment measures are individual subtests of larger batteries designed to measure multiple domains of cognitive functioning; such as, the Judgment Questionnaire subtest of the Neurobehavioral Cognitive Status Exam (NCSE JQ; Northern California Neurobehavioral Group, Inc., 1988), the Problem Solving, and Health and Safety Subscales of the Independent Living Scales (ILS; Loeb, 2006), and the Judgment subtest of the Neuropsychological Assessment Battery (NAB-JDG; Stern & White, 2003). Stand-alone measures of judgment include the Kitchen Picture Test (KPT; Mansbach, MacDougall, Clark, & Mace, 2014), which is a brief screening measure involving a single scenario; The Judgment Assessment Tool (JAT; Escudier et al., 2016), which measures two specific aspects of judgment and has not yet been validated in English, and the Test of Practical Judgment

(TOP-J; Rabin et al., 2007). Of these tests, the NAB-JDG subtest and the TOP-J are both brief interview based measures.

The NAB-JDG subtest is designed to assess judgment and decision making skills related to situations likely to be encountered during daily life (Stern & White, 2003). The authors suggest impaired performance on the NAB-JDG subtest may indicate difficulties with problem solving, poor knowledge of home safety and health or medical issues, and potentially reduced capacity to make decisions in these areas. The TOP-J on the other hand, is designed to assess judgment related to safety, medical, social, ethical, and financial issues and was developed in an effort to address an identified need for an objective, clinically relevant, and psychometrically sound measure for use with older adults (Rabin et al., 2007).

Among the studies to date, research has shown that individuals with AD perform more poorly on both the NAB-JDG subtest and the TOP-J than individuals without AD. For example, Gavett et al. (2012) examined diagnostic accuracy for several NAB subtests and found that individuals with AD performed significantly poorer on the NAB-JDG subtest compared to a heterogeneous sample of individuals without AD (including cognitively healthy, individuals diagnosed with MCI and other non-AD dementias). MacDougall and Mansbach (2013) found that in an assisted living sample, NAB-JDG performance correlated with measures of executive functioning, activities of daily living, and performance on a cognitive screening test. In addition, individuals that were deemed to lack capacity to consent for the study performed more poorly on the judgment measure than individuals who were determined to have capacity to consent (MacDougall & Mansbach, 2013).

On the TOP-J, individuals with AD perform more poorly than individuals with MCI and individuals with MCI perform more poorly than healthy controls, which demonstrates that the TOP-J is sensitive to declining judgment ability in accord with disease severity (Rabin et al., 2007). Performance on the TOP-J is also correlated with performance on other neuropsychological measures of executive functioning such as the WCST (Heaton et al, 1993) and the Letter Fluency subtest of the Delis-Kaplan Executive Functioning System (D-KEFS; Delis, Kaplan, & Kramer, 2001; Rabin et al., 2009) and with gray matter density in prefrontal regions that are associated with executive functioning (Rabin et al., 2009). Measures of reliability are higher for the TOP-J than the NAB-JDG (i.e. test-retest stability of .86 versus .43, inter-rater reliability of .96 versus .85, and internal consistency of .68 versus .45) (Rabin et al., 2007; Stern & White, 2003); however, it is important to note that psychometric properties of the TOP-J were derived from a small sample of 134 individuals.

To date, research comparing performance across measures of judgment is limited and there have been no systematic studies specifically assessing the relationship between performance on the TOP-J and NAB-JDG subtest. As both of these measures are brief and utilize an interview format, the primary aim of the present study is to gain better understanding of the relationship between the TOP-J and the NAB-JDG in an outpatient clinical sample of patients evaluated in a neurodegenerative disease specialty clinic. When administering the judgment measures in clinical practice, a discrepancy between scores on the two scales was often observed by clinicians. Given these observations in clinical practice, it was hypothesized that scores on the TOP-J and the NAB-JDG would have poor strength of

agreement. Understanding the relationship between available measures of judgment such as the TOP-J and NAB-JDG may aid clinicians in selecting appropriate measures and interpreting their results.

Method

Participants

Data were drawn from the archival medical records of patients seen in an outpatient community-based neurology center specializing in diagnosis and treatment of neurodegenerative disease (e.g., AD, MCI, dementia with Lewy-bodies, Frontotemporal dementia, Parkinson's disease, Multiple Sclerosis). Records from a total of 61 consecutive referrals were reviewed and used in subsequent analyses. A cognitive screening measure (described below) was administered during an initial neurology visit and subsequent neuropsychological assessment consisted of completion of a standard battery of cognitive measures including two measures of judgment (described below). The neuropsychological battery was administered by a Certified Specialist in Psychometry, postdoctoral fellow, or advanced graduate student with training in administering and scoring neuropsychological tests under the supervision of a licensed neuropsychologist. The analyzed sample was 52.5% female, predominantly Caucasian (90.2%), with an average age of 73 years ($SD = 9.7$; $range = 50 - 92$ years) and an average of 15.5 years of education ($SD = 2.7$; $range = 11 - 20$ years). Working diagnostic considerations for patients in the present sample included: AD ($n = 15$), MCI ($n = 12$), Vascular dementia (VAD; $n = 6$), dementia with Lewy-bodies (DLB; $n = 3$), Multiple Sclerosis (MS; $n = 2$), Parkinsonian syndromes (PS; $n = 2$), cognitive disorder not otherwise specified (COG NOS; $n = 15$), psychiatric illness including depression and anxiety (PSY; $n = 3$), cognitively normal ($n = 2$), and other neurological conditions including cerebral atrophy (Other; $n = 1$). Diagnoses were determined through consensus conference and neuropsychological assessment results, including both judgment measures, were used to aid in diagnostic formulation. This study was reviewed and approved by the Cleveland Clinic Institutional Review Board.

Measures

Test of Practical Judgment.—The TOP-J is a brief interview based measure designed to assess judgment ability in older adults. The TOP-J consists of 9 open-ended questions related to safety, medical, social/ethical, and financial issues (Rabin et al., 2007). The examiner verbally presents brief scenarios involving hypothetical everyday problems or situations and the examinee is asked to provide aloud their proposed solution or response to each scenario. The examiner records the individual's responses verbatim and ambiguous or unclear responses are queried. To score each item, the examinee's response is compared to a list of sample responses and assigned a score of 0, 1, 2, or 3 points. The score for all 9 items is summed to create a total score. The highest possible score is 27, with higher scores indicating better judgment. Older adults with no cognitive complaints obtained a mean score of 23 ($SD = 2.4$) (Rabin et al., 2007); thus, these values can be used to convert the raw score to a z-score. The Top-J takes approximately 10 minutes to administer and score.

Neuropsychological Assessment Battery, Judgment Subtest.—The NAB-JDG is a brief interview based measure designed to assess judgment and decision making skills related to situations likely to be encountered during daily life (Stern & White, 2003). The NAB-JDG consists of 10 open-ended questions related to home safety, medical, and ethical/social issues. The examiner presents each question verbally and records the examinees response verbatim with unclear or ambiguous responses being queried. Each item is assigned a score of 0, 1, or 2 points and the score for all 10 items is summed to create a total raw score, with the highest possible score being 20. The NAB demographically corrected norm sample obtained a mean score of 14.7 ($SD = 2.9$). NAB normative data (NAB Demographically Corrected Norms Manual) is used to convert the total raw score to a z-score that is not demographically corrected. The z-score can be converted to a demographically adjusted (sex, age, and education) T-score.

Montreal Cognitive Assessment.—The MoCA (Nasreddine et al., 2005) is a brief cognitive screening measure that takes approximately 10 minutes to administer and can be administered by various levels of health care providers. The MoCA consists of 12 individual tasks grouped into cognitive domains including: visuospatial/executive functioning, naming, attention, language, abstraction, memory, and orientation. A total score is calculated and an educational correction is made if applicable (i.e., one point added for individuals with 12 years of education or less). A total score of 26 indicates cognitive impairment. The MoCA has been well validated for use in neurodegenerative disease populations (Julayanont & Nasreddine, 2016), including the current study's clinic population (Vogel, Banks, Cummings, & Miller, 2015).

Data Analysis

Analyses were conducted using IBM Statistics for Windows (Version 22). Because whole numbers are necessary for calculating concordance correlation coefficients (CCC), NAB-JDG and TOP-J z-scores were converted to T-scores using a conversion formula. Demographically adjusted T-scores are not available for the TOP-J, so NAB-JDG T-scores were also not demographically adjusted. CCC and root mean squared differences (RMSD) were calculated based on a *consistency* definition of agreement (Barchard, 2012; Schuck, 2004) between NAB-JDG T-score and TOP-J T-score. In contrast to an absolute definition of agreement (i.e., $Y = X$), a *consistency* definition measures the degree of additivity, or how well a measure can be equated from another (i.e., $Y = X + a$), between scales and ignores differences in the means (Schuck, 2004). The CCC is preferred over the Pearson correlation coefficient when examining the rate of agreement between two measures (Barchard, 2012; King & Chinchilli, 2001; Lin, 1989; Berg et al., 2016). The Pearson correlation coefficient demonstrates whether there is a linear relationship between measures but unlike the CCC, the Pearson correlation coefficient does not take into account differences in means and standard deviations. CCC values range from 0 to 1 and are interpreted similarly to Pearson correlations, with values below 0.90 representing poor agreement (McBride, 2005). The RMSD is an additional indication of agreement and represents the average amount that scores change between measures (Barchard, 2012). RMSD values range from 0 to the difference between the highest and lowest possible scores on the measures. For further

discussion on the CCC and RMSD, and its comparison to the Pearson correlations, the interested reader is referred to Barchard (2012).

Judgment discrepancy scores were calculated by subtracting the NAB-JDG T-score from the TOP-J T-score for each participant. Discrepancy scores were used then to generate a Bland-Altman plot and in a one-sample t test to test the difference from zero (i.e., perfect agreement) to evaluate for systematic bias. A linear regression was also conducted to test for proportional bias, which would indicate that the difference between the TOP-J and NAB-JDG T-scores is a function of the average of these measures. In order to explore whether consistency between the two measures would be different if demographically adjusted NAB-JDG scores were used, these analyses were repeated using demographically corrected NAB-JDG T-scores and non-adjusted Top-J T-scores.

To evaluate the relationship between both judgment tasks and measures of executive functioning, Pearson correlation coefficients were calculated between TOP-J T-scores, NAB-JDG T-Scores and raw scores from the following D-KEFS subtests: 1) Letter Fluency; 2) Trail Making Test – Number-Letter Switching; 3) Color-Word Interference Test Inhibition. Pearson correlation coefficients were also calculated between T-scores from the judgment tasks and the total score from the MoCA as a marker of general cognitive functioning. Discriminating power for each judgment measure was determined by calculating the percent of items correctly solved by the normative groups.

Results

T-scores on the TOP-J ranged from 0 to 58 and NAB-JDG T-scores that were not demographically adjusted ranged from 19 to 75. The CCC between non-adjusted NAB-JDG T-score and TOP-J T-score was 0.39, indicating poor strength of agreement (McBride, 2005). The RMSD was 13.7, suggesting the average change between estimates for a given individual is more than one standard deviation. Discrepancies between T-scores ranged from -48.5 to 16, with a mean discrepancy of -13.7 ($SD = 13.7$). This discrepancy was significantly different from zero ($t(60) = -7.81, p < .001$) indicating the presence of a fixed bias such that the NAB-JDG systematically generates higher scores than the TOP-J. The distribution of discrepancy scores is presented in Figure 1 and shows 55.7% of participants obtained discrepancy scores greater than one standard deviation. Four individuals had discrepancy score greater than 40.0, representing outliers in the present distribution. Further exploration of these four cases revealed that 3 were female, ages ranged from 72–83 years and MoCA scores ranged from 22–27, which is consistent with the total sample. The Bland-Altman plot (Figure 2) does not indicate the presence of proportional bias ($R^2 = .05, F(1, 59) = 2.83, p = .10$; judgment mean $\beta = .21, p = .10$). There was a positive correlation between MoCA total score and TOP-J T-score and MoCA total score and NAB-JDG T-score as depicted in Figure 3. Correlations between the judgment measures and tests of executive functioning are presented in Table 1.

The above analyses were repeated using the NAB-JDG demographically adjusted T-score instead of the non-adjusted T-score. NAB-JDG demographically adjusted T-scores ranged from 19 to 71. The CCC between the adjusted NAB-JDG T-score and TOP-J T-score was

0.32, indicating poor strength of agreement. The RMSD was 14.6 and discrepancies between T-scores ranged from -45 to 22, with a mean discrepancy of -11.9 ($SD = 14.6$). This discrepancy was significantly different from zero ($t(60) = -6.35, p < .001$) indicating the presence of a fixed bias. The distribution of discrepancy scores is presented in Figure 1. The Bland-Altman plot (Figure 2) does not indicate the presence of proportional bias ($R^2 = .19, F(1, 59) = 2.31, p = .13$; judgment mean $\beta = .19, p = .13$). MoCA total scores ranged from 10 to 29 with a mean of 21.6 ($SD = 4.5$), which is considered impaired. The TOP-J normative group obtained 85% of all items correct and the NAB-JDG normative group obtained 73.5% of all items correct.

Discussion

Findings from this study support our hypothesis that the TOP-J and NAB-JDG have poor strength of agreement. In a neurodegenerative disease clinic population, individuals generally scored within normative expectations on the NAB-JDG and typically obtained significantly higher scores on the NAB-JDG than the TOP-J. Both measures of judgment demonstrated a positive relationship with overall cognitive screening as measured by the MoCA and with additional measures of executive functioning. However, measures of agreement between the NAB-JDG and TOP-J showed poor consistency. Even though the TOP-J and NAB-JDG are both designed to measure judgment ability, the NAB-JDG systemically generates scores that are on average, 13.7 points higher than scores on the TOP-J (using NAB-JDG non-adjusted T-scores) and 11.9 points higher than scores on the TOP-J (using NAB-JDG adjusted T-scores); these discrepancies both reflect the presence of a fixed bias, which is present whether NAB-JDG scores are demographically adjusted or unadjusted. No evidence of proportional bias was evident between the two judgment measures. Taken together, results indicate there is a significant amount of inconsistency between these two judgment tests that would contraindicate using the measures interchangeably.

Reviewing the published normative data for the judgment tests indicate that the NAB-JDG has greater discriminating power than the TOP-J (i.e., 73.5% versus 85% of correctly solved items for healthy individuals). Chapman and Chapman (1973) suggest that “the more discriminating test will yield a greater performance deficit than the less discriminating test when given to a group of subjects with lower ability than the control group” (p.384); however, in our sample we observed the opposite such that individuals performed better on the NAB-JDG, which is the test with greater discrimination power, than on the TOP-J. As a result, the difference between discrimination power of the TOP-J and NAB-JDG does not seem to be substantial enough to be responsible for the observed differences between NAB-JDG and TOP-J performance.

While the NAB-JDG and TOP-J are both designed to assess judgment, they may be measuring different aspects of this multifaceted construct. The NAB-JDG mainly consists of questions comprised of short, single sentences related to understanding basic safety and hygiene. The NAB-JDG items mainly ask the individual why they should engage in particular behaviors. In contrast, the TOP-J consists of questions that involve multiple longer sentences comprised of more complex language. The TOP-J items provide the examinee

with problems involving everyday situations (i.e. medical, financial, social/ethical, and safety) and the examinee is asked to generate a potential solution to the hypothetical problem, as opposed to describing the rationale for completing a particular behavior. Thus, the item content of the TOP-J may elicit a more sophisticated or higher level of decision-making and problem solving.

We propose that the NAB-JDG may be a more appropriate choice when assessing individuals with more severe impairments that may have difficulty with the longer and more complex language and scenarios on the TOP-J. The NAB-JDG may also be more appropriate when there are questions about a patient's ability to engage in basic hygiene and self-care behaviors. Given that the TOP-J is sensitive to distinguishing between individuals with MCI versus healthy controls (Rabin et al., 2007), it may be more appropriate for clinicians to select the TOP-J for individuals who may demonstrate more subtle impairments so these impairments are not missed, or for those who are still able to manage very basic activities of daily living (ADL), but may have difficulty with more complex ADLs. Relying solely on the NAB-JDG subtest in a neurodegenerative disease population could result in under identification of individuals with impairments in higher level judgment skills beyond safety and hygiene. When feasible, it may be beneficial for these measures of judgment to be used in conjunction with one another in order to obtain more comprehensive information regarding different components of judgment than what either measure is capable of providing on its own, though ultimately, as is the case with most tests, selection should be based on the clinical needs and purpose of the evaluation. For patients with more advanced disease or greater cognitive impairment, the NAB-JDG may be the more appropriate test, though for patients with more preserved cognition or milder impairments, the TOP-J may be a better choice.

One limitation of the current study is that participants were drawn from a center specializing in a variety of neurodegenerative diseases where neuropsychological evaluation is used to inform differential diagnosis. This not only limits generalizability to other clinical populations, it also precludes refined analysis based on disease. The advantage of the present sample, however, is that the diversity in our referral stream is more likely to reflect that seen in a typical outpatient setting, therefore increasing the generalizability of the present findings. Future directions for research should further explore the appropriate use of each measure and include evaluation of the relationship between the TOP-J and NAB-JDG in specific disease populations (e.g. AD versus MCI). Identifying the clinical characteristics of poor performance on each of the judgment measures, as well as, elucidating the underlying latent constructs (e.g., exploratory factor analysis) being measured could be particularly useful in guiding clinical practice. Another future direction involves investigation of the neuroanatomical correlates of performance on these judgment measures using neuroimaging techniques. A major limitation in this, and most studies on judgment, however, is that there is as of yet, no validated, external criterion against which clinical determinations can be made.

Overall, the present study has demonstrated that in a neurodegenerative disease clinic population, the TOP-J and NAB-JDG provide estimates of judgment ability that are systematically different. Individuals perform significantly better on the NAB-JDG than the

TOP-J, and generally within normative expectation. Thus, clinicians will need to carefully review the item-level content of each scale and decide based on the contextual factors of their clinic and patient population, which measure is best.

Acknowledgments

Funding

Research reported in this publication was partially supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 5P20GM109025.

References

- Barchard KA (2012). Examining the reliability of interval level data using root mean square differences and concordance correlation coefficients. *Psychological Methods*, 17(2), 294–308. doi: 10.1037/a0023351 [PubMed: 21574711]
- Berg JL, Durant J, Banks SJ, & Miller JB (2016). Estimates of premorbid ability in a neurodegenerative disease clinic population: comparing the Test of Premorbid Functioning and the Wide Range Achievement Test, 4th Edition. *The Clinical Neuropsychologist*, 30(4), 647–557. doi: 10.1080/13854046.2016.1186224
- Capucho PHFV & Brucki SMD (2011). Judgment in mild cognitive impairment and alzheimer's disease. *Dementia & Neuropsychologia*, 5(4), 297–302. [PubMed: 29213756]
- Chapman LJ, & Chapman JP (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin*, 79(6), 380–385. [PubMed: 4707457]
- Escudier F, Leveille E, Charbonneau S, Cole J, Hudon C, Bedirian V, & Scherzer P (2016). Evaluating Decision-Making: Validation and Regression-Based Normative Data of the Judgment Assessment Tool. *Archives of Clinical Neuropsychology* doi:10.1093/arclin/acw019
- Gavett BE, Lou KR, Daneshvar DH, Green RC, Jefferson AL, & Stern RA (2012). Diagnostic accuracy statistics for seven Neuropsychological Assessment Battery (NAB) test variables in the diagnosis of Alzheimer's disease. *Applied Neuropsychology: Adult*, 19(2), 108–115. doi: 10.1080/09084282.2011.643947 [PubMed: 23373577]
- Julayanont P, & Nasreddine ZS (2016). Montreal cognitive assessment (MoCA): Concept and clinical review. In Lerner AJ (Ed.), *Cognitive screening instruments: A practical approach* (pp.139–195). Cham, Switzerland: Springer.
- King TS, & Chinchilli VM (2001). A generalized concordance correlation coefficient for continuous and categorical data. *Stat Med*, 20(14), 2131–2147. doi:10.1002/sim.845 [PubMed: 11439426]
- Lin LI (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255–268. [PubMed: 2720055]
- Loeb PA (2006). *Independent living scales Manual* San Antonio, TX: Psychological Corporation.
- Heaton RK, Chelune GJ, Talley JL, Kay CG, Curtiss G *Wisconsin Card Sorting Test Manual* (1993). Odessa, FL: Psychological Assessment Resources.
- MacDougall EE, & Mansbach WE (2013). The Judgment Test of the Neuropsychological Assessment Battery (NAB): psychometric considerations in an assisted-living sample. *The Clinical Neuropsychologist*, 27(5), 827–839. doi:10.1080/13854046.2013.786759 [PubMed: 23570279]
- Mansbach WE, MacDougall EE, Clark KM, & Mace RA (2014). Preliminary investigation of the Kitchen Picture Test (KPT): a new screening test of practical judgment for older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology & Cognition*, 21(6), 674–692. doi:10.1080/13825585.2013.865698
- McBride GB (2005). *A proposal for Strength-of-Agreement Criteria for Lin's Concordance Correlation Coefficient* Hamilton, New Zealand.
- Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, ... Chertkow H (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive

impairment. *J Am Geriatr Soc*, 53(4), 695–699. doi:10.1111/j.1532-5415.2005.53221.x [PubMed: 15817019]

Northern California Neurobehavioral Group. (1988). *Manual for the Neurobehavioral Cognitive Status Exam* Fairfax, CA: Author.

Rabin LA, Borgos MJ, & Saykin AJ (2008). A survey of neuropsychologists' practices and perspectives regarding the assessment of judgment ability. *Applied Neuropsychology*, 15(4), 264–273. doi:10.1080/09084280802325090 [PubMed: 19023743]

Rabin LA, Borgos MJ, Saykin AJ, Wishart HA, Crane PK, Nutter-Upham KE, & Flashman LA (2007). Judgment in older adults: development and psychometric evaluation of the Test of Practical Judgment (TOP-J). *Journal of Clinical and Experimental Neuropsychology*, 29(7), 752–767. doi: 10.1080/13825580601025908 [PubMed: 17896200]

Rabin LA, Saykin AJ, West JD, Borgos MJ, Wishart HA, Nutter-Upham KE, ... Santulli RB (2009). Judgment in Older Adults with Normal Cognition, Cognitive Complaints, MCI, and Mild AD: Relation to Regional Frontal Gray Matter. *Brain Imaging & Behavior*, 3(2), 212–219. doi:10.1007/s11682-009-9063-6 [PubMed: 24741381]

Schuck P (2004). Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Quality of Life Research*, 13(3), 571–586. [PubMed: 15130022]

Stern RA & White T (2003). *Neuropsychological Assessment Battery: Psychometric and technical manual* Lutz, FL: Psychological Assessment Resources.

Strauss E, Sherman EMS, & Spreen O (2006). *A compendium of neuropsychological tests: Administration, norms, & commentary*, third edition NY, NY: Oxford University Press.

Vogal SJ, Banks SJ, Cummings JL, & Miller JB (2015). Concordance of the Montreal cognitive assessment with standard neuropsychological measures. *Alzheimers and Dementia*, 1(3), 289–294.

Wechsler D (1997). *Wechsler adult intelligence scale*, third edition: Administration and scoring manual San Antonio, TX: The Psychological Corporation.

Figure 1.

Frequency distributions of discrepancy scores between the Test of Practical Judgment and Neuropsychological Assessment Battery Judgment subtest non-demographically adjusted T-scores and Test of Practical Judgment and Neuropsychological Assessment Battery Judgment subtest demographically adjusted T-scores. *Note.* TOPJ = Test of Practical Judgment T-score; NAB = Neuropsychological Assessment Battery, Judgment subtest T-score.

Figure 2.

Bland-Altman Plots of the discrepancy between the Test of Practical Judgment and Neuropsychological Assessment Battery Judgment subtest T-scores (non-demographically adjusted and demographically adjusted) as a function of the average within-subject estimate with 95% confidence interval around the mean difference. *Note.* TOPJ = Test of Practical Judgment T-score; NAB = Neuropsychological Assessment Battery, Judgment Subtest.

Figure 3.

Cognitive screening scores and performance on the Test of Practical Judgment and Neuropsychological Assessment Battery Judgment subtest. *Note.* TOPJ = Test of Practical Judgment T-score; NAB = Neuropsychological Assessment Battery, Judgment Subtest; MoCA = Montreal Cognitive Assessment.

Table 1.

Correlations, means, and standard deviations of neuropsychological assessment battery

Measure	N	M	SD	Correlations				
				1	2	3	4	5
1. TOP-J	61	35.11	13.57					
2. NAB-JDG	61	48.83	11.11	.40**				
3. NAB-JDG Adj	61	47.00	11.26	.32*	.96**			
4. FAS	60	31.60	11.40	.47**	.61**	.56**		
5. CWIT Inh	54	79.19	23.22	-.10	-.54**	-.51**	-.56**	
6. TMT4	49	119.0	41.95	-.35*	-.54**	-.49**	-.55**	.68**

*Note.** $p < .05$.** $p < .01$.

TOP-J = Test of Practical Judgment T-score; NAB-JDG = Neuropsychological Assessment Battery, Judgment subtests non-demographically adjusted T-score; NAB-JDG adj = Neuropsychological Assessment Battery, Judgment subtest demographically adjusted T-score; FAS = Letter Fluency total score; CWIT Inhib = Color-Word Interference Test, Inhibition Trial raw score; TMT4 = Trail Making Test, Condition 4 raw score.