Journal of the International Neuropsychological Society (2024), **30**, 350–359
doi:10.1017/S1355617723000723

Research Article

Six elements test vs D-KEFS: what does "Ecological Validity" tell us?

Yana Suchy, Michelle Gereau Mora, Stacey Lipio Brothers and Libby A. DesRuisseaux
Department of Psychology, University of Utah, Salt Lake City, UT, USA

Abstract

Objective: Extensive research shows that tests of executive functioning (EF) predict instrumental activities of daily living (IADLs) but are nevertheless often criticized for having poor ecological validity. The Modified Six Elements Test (MSET) is a pencil-and-paper test that was developed to mimic the demands of daily life, with the assumption that this would result in a more ecologically valid test. Although the MSET has been extensively validated in its ability to capture cognitive deficits in various populations, support for its ability to predict functioning in daily life is mixed. This study aimed to examine the MSET's ability to predict IADLs assessed via three different modalities *relative* to traditional EF measures. **Method:** Participants (93 adults aged 60 – 85) completed the MSET, traditional measures of EF (Delis-Kaplan Executive Function System; D-KEFS), and self-reported and performance-based IADLs in the lab. Participants then completed three weeks of IADL tasks at home, using the Daily Assessment of Independent Living and Executive Skills (DAILIES) protocol. **Results:** The MSET predicted only IADLs completed at home, while the D-KEFS predicted IADLs across all three modalities. Further, the D-KEFS predicted home-based IADLs *beyond* the MSET when pitted against each other, whereas the MSET did not contribute beyond the D-KEFS. **Conclusions:** Traditional EF tests (D-KEFS) appear to be superior to the MSET in predicting IADLs in community-dwelling older adults. The present results argue against replacing traditional measures with the MSET when addressing functional independence of generally high-functioning and cognitive healthy older adult patients.

Keywords: Daily functioning; executive function; ecological validity; aging; activities of daily living

(Received 1 May 2023; final revision 19 August 2023; accepted 13 September 2023; First Published online 11 March 2024)

Introduction

Ecological validity is typically conceptualized as a test's ability to predict various aspects of daily functioning (Long, 1996; Sbordone, 1996), which is often, although by no means exclusively, operationalized as the ability to engage in basic and instrumental activities of daily living (IADLs)[1]. From a neurocognitive standpoint, IADLs rely on executive functioning (EF; for review, see Overdorp et al., 2016), that is, a set of higher-order neurocognitive processes necessary for execution of goal-directed and future-oriented behavior (e.g., Lezak et al., 2012; Suchy, 2015). However, traditional neuropsychological tests of EF have been criticized for having poor ecological validity (e.g., Allain et al., 2014; Chevignard et al., 2008; Jovanovski et al., 2012; Longaud-Valès et al., 2016; Renison et al., 2012; Rosetti et al., 2018; Shimoni et al., 2012; Torralva et al., 2012; Valls-Serrano et al., 2018; Werner et al., 2009). This criticism is surprising, given that such tests have repeatedly shown effectiveness in predicting IADLs (e.g., Bell-McGinty et al., 2002; Boyle et al., 2004; Cahn-Weiner et al., 2002; Johnson et al., 2007; Karzmark et al., 2012; Nguyen et al., 2020; Perna et al., 2012; Putcha & Tremont, 2016; Sudo et al., 2015).

The apparently unwarranted criticism of traditional EF tests is likely related to inconsistent conceptualizations of the term "ecological validity." Specifically, in addition to the notion that tests that predict real-world functioning are ecologically valid, ecological validity is sometimes conceptualized as a combination of *both* the test's ability to predict functioning *and* the test's resemblance to daily life (Franzen & Wilhelm, 1996). Since this latter characteristic is conspicuously lacking in traditional neuropsychological tests, there have been calls for the development of new tests that would resemble the "real world" (e.g., Burgess et al., 2006; Spooner & Pachana, 2006). These calls led to the introduction of many face-valid[2] measures, ranging from paper-and-pencil tests (e.g., Kenworthy et al., 2020; Torralva et al., 2012; Wilson, 1993; Zartman et al., 2013) to tests performed in real (e.g., Shallice & Burgess, 1991), mock (e.g., Chevignard et al., 2010;

[1]We acknowledge that ecological validity can also be considered in relation to a variety of additional everyday outcomes, such as job performance, school performance, driving ability/safety, as well as specific aspects of IADLs, such as the ability to manage medications or finances, to name a few.

[2]In this article, we use "face-valid" to refer to any number of overt test characteristics that are thought to increase similarity with the "real world" and intended to thereby increase the test's "ecological validity." Such characteristics may range from a simple lack of structure and high reliance on multitasking (intended to mimic lack of structure and multitasking in daily life) to highly "naturalistic" demands, such as performance of actual IADL tasks (often cooking or shopping) in mock, virtual, or real settings (e.g., performing a test in an actual kitchen or in an actual supermarket).

Corresponding author: Yana Suchy; Email: yana.suchy@psych.utah.edu
Cite this article: Suchy Y., Gereau Mora M., Brothers S.L., & DesRuisseaux L.A. (2024) Six elements test vs D-KEFS: what does "Ecological Validity" tell us?. *Journal of the International Neuropsychological Society*, **30**: 350–359, https://doi.org/10.1017/S1355617723000723

Lamberts et al., 2010; Rosenblum et al., 2015; Schmitter-Edgecombe et al., 2021), and virtual environments (e.g., Chicchi Giglioli et al., 2021; Josman et al., 2009; Jovanovski et al., 2012). However, translation of these measures into clinical practice has been lacking, with only one such instrument, the Behavioural Assessment of Dysexecutive Syndrome (BADS) battery (Wilson et al., 1998), currently utilized clinically (Rabin et al., 2007). The BADS is comprised of six paper-and-pencil tasks designed to mimic daily life. From among these, the Modified Six Elements Test (MSET) is often considered the most sensitive to cognitive deficits (Burgess et al., 1998; Burgess et al., 2006; Emmanouel et al., 2014).

The MSET is modeled after the Multiple Errands Test (Shallice & Burgess, 1991) and is intended to approximate demands of daily life. It has been shown to detect cognitive deficits in persons with mild cognitive impairment and Alzheimer's dementia (Canali et al., 2007; Espinosa et al., 2009; Esposito et al., 2010; da Costa et al., 2022), Parkinson's disease (Perfetti et al., 2010), brain injury (Emmanouel et al., 2014; Gilboa et al., 2019; Norris & Tate, 2000; Wilson et al., 1998), autism spectrum disorder (Hill & Bird, 2006; White et al., 2009), schizophrenia (Liu et al., 2011; Wilson et al., 1998), and substance use (Valls-Serrano et al., 2018; Verdejo-García & Pérez-García, 2007). However, findings about MSET's ability to predict daily functioning have been mixed. Specifically, while some studies demonstrated associations of the MSET with behavioral measures (Alderman et al., 2003; Chevignard et al., 2008; Conti & Brucki, 2018; Frisch et al., 2012) and rating scales of IADLs or daily EF lapses (Allain et al., 2014; Burgess et al., 1998; Clark et al., 2000; Emmanouel et al., 2014; Jovanovski et al., 2012; Lamberts et al., 2010; Renison et al., 2012; Rochat et al., 2009), others yielded null results (e.g., Bertens et al., 2016; Gilboa et al., 2014; Jovanovski et al., 2012; Norris & Tate, 2000; Romundstad et al., 2022; Roy et al., 2015; Schaeffer et al., 2022). Despite this mixed evidence, MSET is routinely described as being "ecologically valid," implying that its ability to predict daily life has been well documented (e.g., Espinosa et al., 2009; O'Shea et al., 2010; de Almeida et al., 2014; Spitoni et al., 2018; Verdejo-García & Pérez-García, 2007; Wilson et al., 1998).

In summary, while traditional EF tests have a large body of evidence supporting their ability to predict various functional outcomes, they are nevertheless criticized for having poor ecological validity, ostensibly due to their lack of face validity. In contrast, MSET is routinely, if not universally, described as an ecologically valid measure, even though the support for its ability to predict functional outcomes is mixed. In addition, by virtue of being widely deemed ecologically valid, MSET is also deemed to be inherently superior to traditional EF tests (Burgess et al., 2006). The purpose of the present study was two-fold: (1) First, given the inconsistent findings, we aimed to *comprehensively* test the ability of MSET to predict daily functioning, using three different modalities of IADL assessment: self-report, lab-based behavioral assessment, and independent performance at home. (2) Second, given the common impression that tests described as "ecologically valid" are inherently superior to traditional tests of EF, we compared MSET to a traditional EF measure as IADL predictors. To these ends, we administered the MSET and subtests from the Delis-Kaplan Executive Function System (D-KEFS) battery to community-dwelling older adults. Participants also completed Lawton IADL questionnaire (Lawton & Brody, 1969), Timed Instrumental Activities of Daily Living test (TIADL; Owsley et al., 2002), and a three-week protocol of IADL tasks completed independently at home (Brothers & Suchy, 2021; Suchy et al., 2022). Given that we previously showed that face validity in and of itself does not improve a test's ability to predict IADLs (Suchy et al., 2022; Ziemnik & Suchy, 2019), we hypothesized that MSET would predict IADLs in all three modalities, but D-KEFS would account for IADL variance beyond MSET.

Table 1. Characteristics of the sample.

	Mean	SD	Minimum	Maximum
Age	69.31	5.59	60	85
Education	16.37	2.56	11	24
DRS-2 raw total	139.03	4.82	116	144
GDS	5.03	5.09	0	30

Note: N = 93; DRS-2=Dementia Rating Scale Second Edition; GDS = Geriatric Depression Scale; SD = Standard Deviation. For three participants with missing GDS scores, the missing values were replaced with the sample mean.

Method

Participants

Participants were 100 older adults recruited into the DAILIES study examining the impacts of contextual factors on daily functioning (see Brothers & Suchy, 2022). For inclusion, participants needed to be at least 60 years of age, living independently, and, per self-report, not previously diagnosed with dementia, mild cognitive impairment, or other significant neurological disorders (e.g., essential tremor, stroke). Participants were excluded if they self-reported color-blindness, uncorrected hearing or visual impairments that would preclude task performance, less than eight years of formal education, or were not fluent/literate in English. Seven participants were excluded due to missing data on primary variables, for a final sample of 93 participants (69% female). Participants were primarily non-Hispanic White (89%), with 5.4% self-reporting being Hispanic/Latine and 5.4% declining to disclose ethnicity. Additionally, 84% were right-handed, 58% lived with a spouse/partner, and 80% were retired. See Table 1 for additional sample characteristics. Approximately 50 participants from the present sample were included in previous studies (Brothers & Suchy, 2021; Suchy et al., 2022), but MSET was not examined in those studies.

Procedures

Participants were screened over the telephone. Eligible participants completed about four hours of baseline testing, including self-report and cognitive measures used for the larger study. At the end of the testing, participants were given instructions and practice items for the at-home assessment of IADLs. After three weeks of completing at-home IADL tasks, participants returned for debriefing and, if interested, feedback about their overall cognitive/psychiatric functioning. Participants were reimbursed \$10 per hour for the baseline visit, \$20 for the feedback visit, and \$4 for each at-home task. The study was approved by the University of Utah Institutional Review Board and was conducted in accordance with Helsinki Declaration. *P* values < .05, two-tailed, were considered statistically significant.

Measures

Characterizing the sample

To characterize the participants' general cognitive status and depressive symptoms, we used the Dementia Rating Scale-Second Edition (DRS-2; Jurica et al., 2001) and the 30-item version of the Geriatric Depression Scale (GDS; Yesavage, 1988), respectively.

Three participants had missing GDS scores that were replaced with the sample mean.

Modified six elements test (MSET)

The MSET (Wilson et al., 1998) is modeled after the Multiple Errands Test (Shallice & Burgess, 1991), designed to rely on cognitive processes needed in daily life, including meta-tasking, initiation, prospective memory, and self-monitoring. The MSET requires examinees perform six tasks within 10 minutes while adhering to certain rules. The tasks include dictating responses to two story prompts, solving and recording answers to two sets of simple arithmetic problems, and recording answers to two sets of object-naming problems. Examinees are instructed that it is not possible to complete all six tasks within the allotted time but that they should (a) complete at least some portion of each task and (b) avoid completing two tasks of the same type in a row. Thus, examinees must spontaneously switch among tasks in accordance with the rules, while avoiding running out of time. The total score consists of the number of tasks attempted (a maximum of six is possible) minus (a) the number of rule breaks and (b) one point for inefficient use of time (defined as spending more than 271 s on any one task). Total possible scores range from zero to six. Prior research has reported low test-retest reliabilities, ranging from .43 to .48 (Bertens et al., 2016; Jelicic et al., 2001), as is common for many tests of EF (Calamia et al., 2013; Suchy & Brothers, 2022), but high interrater reliability ($r = .88$; Wilson et al., 1998).

D-KEFS. The D-KEFS is a battery of traditional EF tasks with low face validity. We generated a composite from four timed subtests (Trail Making Test, Verbal Fluency, Design Fluency, and Color-Word Interference; Delis et al., 2001), consistent with prior research (e.g., Franchow & Suchy, 2015, 2017). The composite was generated from scores designated as "primary" in the test manual. First, raw subtest scores were converted to scaled scores based on norms[3] (Delis et al., 2001). Next, we generated a single score for each subtest by averaging across the scores from the relevant executive conditions within that subtest: one condition of the Trail Making Test (number-letter switching completion time), three Design Fluency conditions (number correct in filled dots, empty dots, and switching), three Verbal Fluency conditions (number correct in letter, category, and category switching), and two Color-Word Interference conditions (interference and interference-switching completion times). We then averaged across the four subtests to generate the final D-KEFS composite. Cronbach's alpha in this sample was .78. Test-retest reliabilities were not tested in the present sample but were previously reported at .90 (Suchy & Brothers, 2022).

Because performance on timed EF measures is influenced by lower-order processes (e.g., graphomotor speed, visual scanning, etc.; Suchy, 2015; Stuss & Knight, 2002), we also generated a lower-order process composite. Specifically, we averaged the scaled scores[3] of subtest conditions designed to isolate lower-order processes as defined by the D-KEFS manual (Delis et al., 2001): four Trail Making Test conditions (visual scanning, number sequencing, letter sequencing, and motor speed) and two Color-Word Interference conditions (color naming and word reading completion time). This composite was used as a covariate to help isolate the EF construct, as done in prior research (e.g., Franchow & Suchy, 2015, 2017; Kraybill & Suchy, 2011; Kraybill et al., 2013). Cronbach's alpha in this sample was .76. Because of the heavy speed demands of these tasks, we refer to this variable as "Processing Speed" below.

Self-reported IADLs

Self-reported IADLs were assessed using the Lawton IADL scale (Lawton & Brody, 1969). Individuals rate their level of independence (on a three-point scale) in seven IADL domains. The scale has been extensively validated (e.g., Mariani et al., 2008; Ng et al., 2006), with a test-retest reliability reported at .85 (Lawton & Brody, 1969). Internal consistency could not be calculated in this sample, as some items lacked variability, as can be expected in high functioning samples. Higher scores on this scale indicate fewer problems. Hereafter, we call this variable "IADLs-Report."

Lab-based IADLs

Participants completed the performance-based Timed Instrumental Activities of Daily Living (TIADL; Owsley et al., 2002), comprised of tasks related to communication (e.g., finding a telephone number in a phone book), finance (e.g., making change), food (reading ingredients on cans of food), shopping (e.g., finding food items on a shelf), and medication management (e.g., reading instructions on medicine bottles). Completion times for the five tasks were converted to z-scores based on the current sample, then averaged to create a speed composite. Errors across the five tasks were summed, then also converted to a z-score based on the current sample. The speed and error composites were then averaged to generate an overall performance score for the TIADL. Cronbach's alpha in this sample was .73. Test-retest reliability was not available for this sample, but was previously reported at .85 (Owsley et al., 2002). Higher scores on this composite are indicative of poorer performance (i.e., more time spent and/or a greater number of errors). Hereafter we call this composite "IADLs-Lab."

Home-based IADLs

To assess participants' IADLs at home, we used the Daily Assessment of Independent Living and Executive Skills (DAILIES) protocol (Brothers & Suchy, 2022). The DAILIES asks participants to complete brief tasks that resemble typical IADLs (e.g., paying utility bills, canceling a doctor's appointment, filling out a rebate form, etc.) six days a week for three weeks. Participants must complete the tasks during specified timeframes (e.g., 9:00 to 11:00 AM) that vary each day to resemble real-world demands, and communicate about task completion with the researchers via email, telephone, or postal mail, (again varied daily to mimic typical real-life demands). Tasks are scored based on timeliness (one point if a response is provided on the correct day, and one point if the response is provided during the allotted timeframe, for a total of two possible points) and accuracy (scores ranging from one to seven depending on complexity). The scores from each task are summed to generate the total DAILIES score (possible maximum of 93 points). Higher scores indicate better performance. Internal consistency was not calculated, as IADLs-home is a "formative" variable intended to provide a sum total of correctly completed tasks during the given timeframe. This is in contrast to "reflective" variables, which are intended to "reflect" a

[3]The D-KEFS raw scores in this study were converted to scaled scores using the normative reference group for adults aged 60-69 years. By doing so, D-KEFS scores could be standardized and combined into a single composite *without correcting for age*. The 60-69-year-old age band was selected because the scores within this age band encompass the widest range of raw scores (as compared to other age bands) and would therefore have the highest probability of avoiding floor or ceiling effects (Delis et al., 2001). We chose this approach to avoid inappropriate mixing of age-corrected and non-age-corrected variables in analyses, which would result in uneven impact of age on various associations among variables, complicating interpretation. We used a similar procedure in other prior studies (DesRuisseaux et al., 2022; Suchy, Brothers, et al., 2020; Suchy, Mullen, et al., 2020).

https://doi.org/10.1017/S1355617723000723 Published online by Cambridge University Press

Table 2. Descriptive statistics of primary dependent and independent variables used in analyses.

Variable name	Mean	SD	Minimum	Maximum
IADLs-Home (Winsorized)	69.49	7.72	45	82
IADLs-Lab (Winsorized)	−0.02	0.84	−1.28	2.14
IADLs-Report (log transformed)	0.60	0.12	0.48	0.85
MSET (log transformed)	0.12	.20	0	0.85
D-KEFS	11.97	1.85	5.63	16.50

Note: $N = 93$. For variables that were normalized via transformation or log-transformation, the transformed scores are presented in the table, as indicated in variable names. D-KEFS = Delis-Kaplan Executive Function System composite score; IADLs-Home = Daily Assessment of Independent Living and Executive Skills (DAILIES) total score; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score; IADLs-Report = Lawton Instrumental Activities of Daily Living raw score; MSET = Modified Six Elements Test.

construct (Kievit et al., 2011). Hereafter, we call this variable "IADLs-Home."

Results

Preliminary analyses

Score distribution

All variables were examined for outliers and normality. IADLs-Home had one outlier and IADLs-Lab had two outliers, which was remedied via Winsorization. IADLs-Report, MSET, GDS, and DRS-2 exhibited a skew that was remedied via log-transform. After these procedures, all variables were normally distributed (all Skewness values <1), except for MSET, which still evidenced slight skew (Skewness = 1.53). Thus, we conducted supplementary non-parametric analyses with the MSET.

Debriefing

Debriefing forms were available for 81 participants. The majority of participants (91.3%) felt the DAILIES tasks were similar to typical tasks they complete in daily life (i.e., responding 'agree' or 'strongly agree' to this item).

Descriptives and zero-order correlations

Descriptives for all dependent and independent variables are presented in Table 2, and zero-order correlations of dependent and independent variables with potential confounds are presented in Table 3. As seen, age, education, general cognitive status, and processing speed were all associated with at least some of the dependent or independent variables. Additionally, we examined correlations among the three IADL variables. Interestingly, while IADLs-Lab and IADLs-Report were correlated ($p = .023$), IADLs-Home was unrelated with lab-based and self-reported IADLs (p-values > .200). The three IADL variables were thus examined individually in all analyses.

Principal analyses

Univariate associations

Zero-order correlations between the dependent and independent variables are presented in Table 4. As seen, D-KEFS was associated with all three IADL variables; contrary to expectation, MSET was associated only with IADLs-Home.

Pitting D-KEFS and MSET against each other

Because both D-KEFS and MSET showed univariate associations with IADLs-Home, we wanted to examine whether these variables predicted IADLs-Home beyond each other. Additionally, even though MSET was unrelated to IADLs-Report and IADLs-Lab, we nevertheless wanted to examine whether D-KEFS predicted these variables beyond MSET. Thus, we ran three general linear regressions, using IADLs-Home, IADLs-Report, and IADLs-Lab as dependent variables and MSET and D-KEFS as predictors. As seen in Table 5, D-KEFS predicted all three IADLs variables beyond MSET, whereas MSET did not contribute variance beyond D-KEFS.

Effects of covariates

We next examined whether D-KEFS still predicted the IADL variables when confounds identified in Table 3 were included as covariates. We ran three general linear regressions, again using the three IADL variables as dependent variables, D-KEFS as a predictor, and age, education, GDS, DRS, and Processing Speed as covariates. As seen in Table 6, D-KEFS again emerged as a unique predictor of IADLs across all three modalities.

Supplementary analyses

Individual D-KEFS subtests

Because the MSET variable was based on a single test, whereas the D-KEFS was a composite of four subtests, one could argue that D-KEFS had an unfair advantage over MSET due to a broader range of sampled processes and higher reliability. To address this issue, we examined partial correlations of the four individual D-KEFS subtests with the three IADL variables, controlling for MSET. As seen in Table 7, all but three correlations were statistically significant, illustrating that even single traditional EF tests with narrower scope and lower reliabilities tend to outperform the MSET.

Homogenizing the sample

In the present sample, two participants' DRS-2 scores fell below 123, the level that is considered normal (Jurica et al., 2001). To ensure that the results were not driven by these two participants, we reran all principal analyses with these two participants removed. The correlation between MSET and IADLs-Home was no longer significant, Spearman's Rho = −.201, $p = .055$. In contrast, D-KEFS maintained all significant results reported in prior analyses (all p values < .05).

Discussion

The aims of the present study were to empirically examine the widely-held assumptions that MSET performance predicts daily IADL functioning and that MSET's clinical utility is superior to that of traditional EF measures. To these ends, we administered the MSET, four subtests from the D-KEFS battery, and three measures of IADLs to a sample of 93 community-dwelling older adults. IADLs were assessed via three modalities: self-report, lab-based behavioral tasks, and home-based tasks completed over three weeks. The key findings are that (a) MSET predicted performance of IADL tasks at home, (b) D-KEFS was associated with IADLs in all three assessment modalities, and (c) D-KEFS accounted for variance in IADLs beyond MSET, as well as beyond potential demographic, cognitive, and psychiatric confounds, whereas MSET did not contribute beyond the D-KEFS.

MSET and ecological validity

The present results are consistent with prior research in that they provide somewhat equivocal, or "soft," evidence of MSET's ability

https://doi.org/10.1017/S1355617723000723 Published online by Cambridge University Press

Table 3. Zero-order correlations of the primary dependent and independent variables with sample characteristics.

	IADLs-Home (Winsorized)	IADLs-Lab (Winsorized)	IADLs-Report (log transformed)	MSET (log transformed)	D-KEFS
Age	−.118	**.291****	−.022	−.045 (−.036)	**−.425*****
Education	.161	−.157	−.119	−.073 (−.106)	**.274****
Sex	.016	−.050	**.222***	−.063 (−.038)	.012
GDS (log transformed)	−.074	.082	**.236***	.047 (.029)	−.119
DRS-2 (log transformed)	**−.270***	**.276***	−.035	.160 (.123)	**−.476*****
Processing Speed	**.245***	**−.391*****	−.075	−.154 (−.181)	**.753*****

Note: $N = 93$. For variables that were normalized via transformation or log-transformation, the normalized scores were used in analyses, as indicated in variable names. DRS-2 = Dementia Rating Scale, Second Edition, raw score; GDS = Geriatric Depression Scale; D-KEFS = Delis-Kaplan Executive Function System composite score; IADL-Report = Lawton Instrumental Activities of Daily Living raw score; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score; IADLs-Home = Daily Assessment of Independent Living and Executive Skills (DAILIES) total score; MSET = Modified Six Elements Test. Non-parametric correlations (Spearman's rho) for the MSET, which exhibited a slight skew, are presented in parentheses. Sex was coded 1 = female, 0 = male (thus, women reported fewer IADL problems on self-report).
*$p < .05$; ** $p < .01$, *** $p < .001$.

Table 4. Zero-order correlations between the primary dependent and independent variables.

	MSET (log transformed)	D-KEFS	Processing speed
IADLs-Home (Winsorized)	**−.220* (−.229*)**	**.378****	**.245***
IADLs-Lab (Winsorized)	.116 (.114)	**−.510****	**−.391****
IADLs-Report (log transformed)	−.122 (−.127)	**−.269****	−.075

Note: $N = 93$. For variables that were normalized via transformation or log-transformation, the normalized scores were used in analyses, as indicated in variable names. D-KEFS = Delis-Kaplan Executive Function System composite score; IADL-Report = Lawton Instrumental Activities of Daily Living raw score; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score; IADLs-Home = Daily Assessment of Independent Living and Executive Skills (DAILIES) total score; MSET = Modified Six Elements Test. Non-parametric correlations (Spearman's rho) for the MSET, which exhibited a slight skew, are presented in parentheses.
*$p < .05$; **$p < .01$.

Table 5. General linear regressions pitting the D-KEFS against the MSET as predictors of instrumental activities of daily living (IADLs).

Dependent variable	Predictors	*B*	*Standard error*	*Beta*	*t*	*p*
IADLs-Home (Winsorized)	Constant	52.76	5.033		10.48	<.001
	MSET	−6.137	3.823	−.157	1.61	.112
	D-KEFS	1.458	.408	.350	3.58	<.001
IADLs-Lab (Winsorized)	Constant	2.705	.541		5.26	
	MSET	.104	.390	.025	.27	.270
	D-KEFS	−.229	.042	−.506	5.49	<.001
IADLs-Report (log transformed)	Constant	.840	.078		44.01	<.001
	MSET	−.103	.059	−.176	1.74	.087
	D-KEFS	−.019	.006	−.300	2.96	.004

Note: $N = 93$. IADL variables used in analyses were normalized as indicated in variable names. MSET = Modified Six Elements Test (log transformed variable was used in analyses); D-KEFS = Delis-Kaplan Executive Function System composite score; IADLs-Home = Home-based performance of IADLs; IADLs-Report = Lawton Instrumental Activities of Daily Living; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score. In corresponding hierarchical models for IADLs-Home, IADLs-Lab, and IADLs-Report, the D-KEFS accounted for 12%, 25%, and 9% of variance beyond the MSET, respectively.

to predict daily functioning. Specifically, while the MSET did predict how participants performed IADLs at home, it was *not* associated with either of the other two IADL measures. Since lab-based and self-reported IADL measures were not related to IADLs performed at home, it is likely that they reflected different aspects of functioning, suggesting that MSET may only tap into a subset of IADL capacity. For example, the home-based IADL protocol was less structured and required greater use of prospective memory than the other IADL measures; similarly, MSET is intended to be less structured and rely more heavily on prospective memory, possibly explaining its association with the home-based IADLs and the lack of association with the other two IADL measures.

Importantly, contrary to the widely-held beliefs about the superiority of tests with high face validity, MSET did not evidence any advantage over D-KEFS. Instead, D-KEFS predicted IADLs well *beyond* MSET. It is thus likely that D-KEFS taps into a broader range of EF processes than MSET. Indeed, traditional EF tests have been shown to predict occupational functioning (for reviews see Gilbert & Marwaha, 2013; Ownsworth & McKenna, 2004), whereas MSET has not (Moriyama et al., 2002), further suggesting that MSET may tap a narrower range of processes. While it could be argued that our D-KEFS composite understandably taps a broader range of processes due to being based on four different subtests, it is noteworthy that D-KEFS subtests outperformed the MSET even when examined individually. Lastly, given that MSET was no longer associated with IADLs once two participants with mildly impaired cognition were removed, it appears that MSET is vulnerable to ceiling effects and as such is not sensitive to subtle deficits. Together, the test's somewhat narrow range of sensitivity, combined with potentially a somewhat narrow scope of IADL capacities to which it is related, may explain the inconsistent findings in prior research.

Alternatively, prior methodological limitations may also explain the inconsistent findings in prior literature. Specifically, prior ecological validations of the MSET reviewed in the introduction utilized only between 24 and 120 participants (median = 47.5). Since about one half of the reviewed studies attempted MSET validation on samples smaller than 50, their

Table 6. General linear regressions predicting three IADL variables, controlling for covariates.

Dependent variable	Predictor	*b*	*Standard error*	*β*	*t*	*p*
IADLs-Home (Winsorized)	Constant	52.179	17.243		3.026	.003
	Age	.017	.173	.012	.098	.922
	Education	.299	.375	.091	.799	.427
	GDS (log transformed)	−.543	2.443	−.024	−.222	.825
	DRS-2 (log transformed)	−3.737	3.956	−.114	−.945	.348
	Processing speed	−.415	.943	−.074	−.440	.661
	D-KEFS	1.624	.813	.382	1.998	.049
IADLs-Lab (Winsorized)	Constant	1.347	1.752		.769	.444
	Age	.015	.018	.102	.866	.389
	Education	−.008	.038	−.024	−.223	.825
	GDS (log transformed)	.128	.248	.054	.517	.607
	DRS-2 (log transformed)	.151	.402	.043	.375	.709
	Processing speed	−.034	.096	−.057	−.359	.720
	D-KEFS	−.174	.083	−.383	−2.108	.038
IADLs-Report (log transformed)	Constant	.916	.253		3.615	<.001
	Age	−.002	.003	−.094	−.749	.456
	Education	.003	.006	.068	.592	.556
	GDS (log transformed)	.048	.036	.146	1.325	.189
	DRS-2 (log transformed)	−.087	.058	−.184	−1.504	.137
	Processing speed	.021	.014	.259	1.530	.130
	D-KEFS	−.036	.012	−.578	−2.995	.004

Note: N = 93. IADL variables used in analyses were normalized as indicated in variable names. DRS-2 = Dementia Rating Scale, Second Edition, raw score; GDS = Geriatric Depression Scale; D-KEFS = Delis-Kaplan Executive Function System composite score; IADLs-Report = Lawton Instrumental Activities of Daily Living; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score; IADLs-Home = Daily Assessment of Independent Living and Executive Skills (DAILIES) total score.

Table 7. Partial correlations between individuals D-KEFS variables and the three IADL variables, controlling for MSET.

	IADLs-Home (Winsorized)	IADLs-Lab (Winsorized)	IADLs-Report (log transformed)
D-KEFS DF	**.233***	**−.348*****	−.167
D-KEFS CWI	.150	**−.414*****	**−.289****
D-KEFS TMT	**.286****	**−.456*****	**−.258***
D-KEFS VF	**.243***	**−.241***	−.132

Note: *N* = 93; *N* = 90 for GDS. IADL variables used in analyses were normalized as indicated in variable names. D-KEFS = Delis-Kaplan Executive Function System; DF = Design Fluency; CWI = Color-Word Interference; TMT = Trail Making Test; VF = Verbal Fluency; IADLs-Report = Lawton Instrumental Activities of Daily Living; IADLs-Lab = Timed Instrumental Activities of Daily Living (TIADLs) total score; IADLs-Home = Daily Assessment of Independent Living and Executive Skills (DAILIES) total score; MSET = Modified Six Elements Test.
*$p < .05$; **$p < .01$; ***$p < .001$.

results may be unstable (e.g., Harris, 1985; Van Voorhis & Morgan, 2007) and vulnerable to non-replication. Poor reliability is yet another possible explanation. Regardless of the sources of inconsistency, the present study suggests that MSET does not incrementally improve upon D-KEFS in predicting functional outcomes, at least not among community-dwelling older adults.

The importance of outcome variables

Past research examining the associations between EF tests and daily functioning has been criticized for relying predominantly on participants or collateral reports about IADLs (for review, see Robertson & Schmitter-Edgecombe, 2016). In contrast, the present study utilized the DALIES, which (per participant endorsement) closely mimics typical daily tasks. The DAILIES has several advantages over typically used methods. First, it reflects IADL performance within the context of daily life, with participants completing the DALIES while also attending to other daily demands, responsibilities, or distractions. Thus, participants needed to independently plan and problem-solve how to interleave the DAILIES within their daily routines while also engaging prospective memory to complete the tasks during the correct time frames.

Second, the DALIES assesses participants' performance of IADLs over a somewhat *extended* period, unlike typical behavioral assessments that examine a single "snapshot" in time. An extended assessment period is critical since EF is known to fluctuate due to a variety of contextual factors (Berryman et al., 2014; Suchy et al., 2022; Franchow & Suchy, 2015, 2017; Tinajero et al., 2018), leading to lapses in IADLs that are *intermittent* and thus cannot be readily captured in a single assessment session. Importantly, since individuals with even mild EF weaknesses are more *vulnerable* to experiencing such fluctuations (Killgore et al., 2009; Williams et al., 2009), predictors of daily functioning need to be sensitive to such subtle EF weaknesses.

Lastly, the DAILIES allowed us to examine whether our tests can predict IADLs *prospectively*, generalizing from performance assessed at one timepoint to a future behavior at home. In contrast, most research examines *concurrent* associations between EF measures and IADL tasks (e.g., Alderman et al., 2003; Conti & Brucki, 2018; Frisch et al., 2012; Suchy et al., 2019), potentially confounding results with a third variable shared in space and time, such as experiencing pain (Attridge et al., 2015; Heyer et al., 2000) or not having slept well the night before testing (Fortier-Brochu et al., 2012; Holding et al., 2021; Miyata et al., 2013). Indeed, such contextual factors have an impact on both EF (Berryman et al., 2014; Niermeyer & Suchy, 2020; Tinajero et al., 2018) and IADLs (Hicks et al., 2008; Stamm et al., 2016; Webb et al., 2018), potentially confounding concurrently observed associations.

Traditional tests of EF, ecological validity, and a call to action

Despite the fact that the present results offer only a somewhat "soft" support of the MSET's ability to predict daily functioning, they nevertheless do, at least technically, support the MSET's ecological validity in that the MSET does possess face validity and does relate (albeit weakly) to daily IADL performance. Interpretation is less straightforward for the D-KEFS. On the one hand, if we define ecological validity as the test's ability to predict functional outcome, then the D-KEFS certainly appears to

https://doi.org/10.1017/S1355617723000723 Published online by Cambridge University Press

be more ecologically valid than the MSET. On the other hand, if we define ecological validity as requiring that the test have face validity, then the D-KEFS cannot be deemed ecologically valid regardless of how well it predicts daily functioning. This latter perspective defies any clinical utility of the term ecological validity. It is our position that the term ecological validity "muddies the waters," misleading clinicians and mischaracterizing the clinical utility of tests. The usage of the term often leads to the erroneous impressions that (a) traditional EF tests cannot possibly predict daily functioning since they lack face validity, and (b) tests with high face validity are inherently able to predict daily functioning and as such are superior to traditional measures. The term ecological validity has been criticized for similar reasons in other areas of psychology as well (Holleman et al., 2020; Kihlstrom, 2021). We therefore call on our field to retire the term ecological validity in favor of more concrete terminology and/or concrete descriptions of what a given test can accomplish in a given population. Indeed, depending on the specific study design, the terms predictive, criterion, and concurrent validity communicate clearly what a given test can or cannot accomplish, thereby being more informative and useful, in both clinical and research contexts.

Limitations

The present study needs to be interpreted within the context of some limitations. First, the sample was predominantly non-Hispanic White, highly educated, and comprised of individuals who were high functioning and cognitively healthy, which may have affected the results. Indeed, MSET was skewed, suffering from a ceiling effect, and the MSET results were driven by two mildly impaired participants. Thus, while it appears that D-KEFS is more sensitive to *subtle* EF deficits than MSET, it is not known whether MSET would outperform the D-KEFS in another, more impaired sample. Additionally, it is unclear whether cultural or linguistic factors would impact performances on currently employed measures unevenly, further impacting results. Thus, we must remind ourselves that validity is specific not only to a given test, but also to a population in which validation occurred.

Second, the present study pitted the D-KEFS composite against a single test. It is possible that a composite of all BADS subtests would perform better than MSET alone, and possibly even better than the D-KEFS. Relatedly, it is possible that the weakness of MSET relative to D-KEFS stems not from its poorer ability to tap into relevant neurocognitive processes (i.e., the measure's content), but rather its poorer psychometric properties, namely poorer reliability. Future research should examine these questions. Meanwhile, although the present results *technically* support ecological validity of MSET, they do *not* support its usage in place of, or in addition to, traditional EF measures.

Conclusions

The present study offers some weak support for the predictive validity of the MSET. However, this support is considerably tempered by the fact that D-KEFS accounted for variance in IADLs beyond MSET, while MSET failed to contribute incrementally to the prediction. Additionally, while D-KEFS was related to two other measures of IADLs (self-report and lab-based performance), MSET was not related to either. Thus, at least among community-dwelling older adults, D-KEFS proves to have a greater clinical utility than MSET. Despite these findings, which favor the D-KEFS over the MSET, the term "ecologically validity" can be applied more confidently to the MSET than to the D-KEFS, due to MSET's greater face validity. These conclusions demonstrate the lack of clinical utility of the term ecological validity. Thus, we argue that "ecological validity" should be avoided in assessment contexts and, as appropriate, replaced with more descriptive terms such as criterion, predictive, or concurrent validity

Funding statement. The study was funded by the senior author's development fund awarded by the University of Utah.

Competing interests. None to declare.

References

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, *9*, 31–44.
Allain, P., Alexandra Foloppe, D., Besnard, J., Yamaguchi, T., Etcharry-Bouyx, F., Le Gall, D., & Richard, P. (2014). Detecting everyday action deficits in alzheimer's disease using a nonimmersive virtual reality kitchen. *Journal of the International Neuropsychological Society*, *20*, 468–477. https://doi.org/10.1017/S1355617714000344
Attridge, N., Noonan, D., Eccleston, C., & Keogh, E. (2015). The disruptive effects of pain on n-back task performance in a large general population sample. *Pain*, *156*, 1885–1891. https://doi.org/10.1097/j.pain.0000000000000245
Bell-McGinty, S., Podell, K., Franzen, M., Baird, A. D., & Williams, M. J. (2002). Standard measures of executive function in predicting instrumental activities of daily living in older adults. *International Journal of Geriatric Psychiatry*, *17*, 828–834. https://doi.org/10.1002/gps.646
Berryman, C., Stanton, T. R., Bowering, K. J., Tabor, A., McFarlane, A., & Moseley, G. L. (2014). Do people with chronic pain have impaired executive function? A meta-analytical review. *Clinical Psychology Review*, *34*, 563–579. https://doi.org/10.1016/j.cpr.2014.08.003
Bertens, D., Fasotti, L., Egger, J. I. M., Boelen, D. H. E., & Kessels, R. P. C. (2016). Reliability of an adapted version of the modified six elements test as a measure of executive function. *Applied Neuropsychology: Adult*, *23*, 35–42. https://doi.org/10.1080/23279095.2015.1012258
Boyle, P. A., Paul, R. H., Moser, D. J., & Cohen, R. A. (2004). Executive impairments predict functional declines in vascular dementia. *The Clinical Neuropsychologist*, *18*, 75–82. https://doi.org/10.1080/13854040490507172
Brothers, S. L., & Suchy, Y. (2022). Daily assessment of executive functioning and expressive suppression predict daily functioning among community-dwelling older adults. *Journal of the International Neuropsychological Society*, *28*, 974–983. https://doi.org/10.1017/S1355617721001156
Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*, 547–558.
Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M.-A., Dawson, D. R., Anderson A. D., Gilbert S. J., Dumontheil I., &Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, *12*, 194–209. https://doi.org/10.1017/S1355617706060310
Cahn-Weiner, D. A., Boyle, P. A., & Malloy, P. F. (2002). Tests of executive function predict instrumental activities of daily living in community-dwelling older individuals. *Applied Neuropsychology*, *9*, 187–191.
Calamia, M., Markon, K., & Tranel, D. (2013). The robust reliability of neuropsychological measures: Meta-analyses of test-retest correlations. *The Clinical Neuropsychologist*, *27*, 1077–1105. https://doi.org/10.1080/13854046.2013.809795
Canali, F., Dozzi Brucki, S. M., & Amodeo Bueno, O. F. (2007). Behavioural assessment of the dysexecutive syndrome (BADS) in healthy elders and alzheimer's disease patients: Preliminary study. *Dementia & Neuropsychologia*, *1*, 154–160. https://doi.org/10.1590/s1980-57642008dn10200007
Chevignard, M. P., Catroppa, C., Galvin, J., Anderson, V. (2010). Development and evaluation of an ecological task to assess executive functioning post childhood TBI: The children's cooking task. *Brain Impairment*, *11*, 125–143. https://doi.org/10.1375/brim.11.2.125

Chevignard, M. P., Taillefer, C., Picq, C., Poncet, F., Noulhiane, M., & Pradat-Diehl, P. (2008). Ecological assessment of the dysexecutive syndrome using execution of a cooking task. *Neuropsychological Rehabilitation*, *18*, 461–485. https://doi.org/10.1080/09602010701643472

Chicchi Giglioli, I. A., Pérez Gálvez, B., Gil Granados, A., & Alcañiz Raya, M. (2021). The virtual cooking task: A preliminary comparison between neuropsychological and ecological virtual reality tests to assess executive functions alterations in patients affected by alcohol use disorder. *cyberpsychology, behavior, and social networking. Cyberpsychology, Behavior, and Social Networking*, *24*, 673–682. https://doi.org/10.1089/cyber.2020.0560

Clark, C., Prior, M., & Kinsella, G. J. (2000). Do executive function deficits differentiate between adolescents with ADHD and oppositional defiant/conduct disorder? A neuropsychological study using the six elements test and hayling sentence completion test. *Journal of Abnormal Child Psychology*, *28*, 403–414.

Conti, J., & Brucki, S. M. D. (2018). Executive function performance test: Transcultural adaptation, evaluation of psychometric properties in Brazil. *Arquivos de Neuro-Psiquiatria*, *76*, 767–774. https://doi.org/10.1590/0004-282x20180127

da Costa, R. Q. M., Pompeu, J. E., Moretto, E., Silva, J. M., dos Santos, M. D., Nitrini, R., & Brucki, S. M. D. (2022). Two immersive virtual reality tasks for the assessment of spatial orientation in older adults with and without cognitive impairment: Concurrent validity, group comparison, and accuracy results. *Journal of the International Neuropsychological Society*, *28*, 460–472. https://doi.org/10.1017/S1355617721000655

de Almeida, R., Macedo, G., Lopes, E., & Monteiro, L. (2014). BADS-C instrument: An ecological perspective of the executive functions in children with attention deficit hyperactivity disorder. *Acta Neuropsychologica*, *12*, 293–303. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2014-45117-002&site=ehost-live

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan Executive Function System (DKEFS): Technical Manual*. The Psychological Corporation.

Emmanouel, A., Mouza, E., Kessels, R. P. C., & Fasotti, L. (2014). Validity of the dysexecutive questionnaire (DEX) ratings by patients with brain injury and their therapists. *Brain Injury*, *28*, 1581–1589. https://doi.org/10.3109/02699052.2014.942371

Espinosa, A., Alegret, M., Boada, M., Vinyes, G., Valero, S., Martínez-Lage, P., Peña-Casanova, J., Becker, J. T., Wilson B. A., & Tárraga, L. (2009). Ecological assessment of executive functions in mild cognitive impairment and mild alzheimer's disease. *Journal of the International Neuropsychological Society*, *15*, 751–757. https://doi.org/10.1017/S135561770999035X

Esposito, F., Rochat, L., Van der Linden, A.-C. J., Lekeu, F., Quittre, A., Charnallet, A., & Van der Linden, M. (2010). Apathy and executive dysfunction in alzheimer disease. *Alzheimer Disease and Associated Disorders*, *24*, 131–137.

Fortier-Brochu, É., Beaulieu-Bonneau, S., Ivers, H., & Morin, C. M. (2012). Insomnia and daytime cognitive performance: A meta-analysis. *Sleep Medicine Reviews*, *16*, 83–94. https://doi.org/10.1016/j.smrv.2011.03.008

Franchow, E. I., & Suchy, Y. (2015). Naturally-occurring expressive suppression in daily life depletes executive functioning. *Emotion*, *15*, 78–89. https://doi.org/10.1037/emo0000013

Franchow, E. I., & Suchy, Y. (2017). Expressive suppression depletes executive functioning in older adulthood. *Journal of the International Neuropsychological Society*, *23*, 341–351. https://doi.org/10.1017/S1355617717000054

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 91–112). Gr Press/St Lucie Press, Inc.

Frisch, S., Förstl, S., Legler, A., Schöpe, S., & Goebel, H. (2012). The interleaving of actions in everyday life multitasking demands. *Journal of Neuropsychology*, *6*, 257–269. https://doi.org/10.1111/j.1748-6653.2012.02026.x

Gilbert, E., & Marwaha, S. (2013). Predictors of employment in bipolar disorder: A systematic review. *Journal of Affective Disorders*, *145*, 156–164. https://doi.org/10.1016/j.jad.2012.07.009

Gilboa, Y., Jansari, A., Kerrouche, B., Uçak, E., Tiberghien, A., Benkhaled, O., Aligon, D., Mariller, A., Verdier, V., Mintegui, A., Abada, G., Canizares, C., Goldstein, A., & Chevignard, M. (2019). Assessment of executive functions in children and adolescents with acquired brain injury (ABI) using a novel complex multi-tasking computerised task: The jansari assessment of executive functions for children (JEF-C©). *Neuropsychological Rehabilitation*, *29*, 1359–1382. https://doi.org/10.1080/09602011.2017.1411819

Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadef, H., & Josman, N. (2014). Is there a relationship between executive functions and academic success in children with neurofibromatosis type 1? *Neuropsychological Rehabilitation*, *24*, 918–935. https://doi.org/10.1080/09602011.2014.920262

Harris, L. N. (1985). *The rationale of reliability prediction*. Wiley Online Library. https://doi.org/10.1002/qre.4680010205

Heyer, E. J., Sharma, R., Winfree, C. J., Mocco, J., Mahon, D. J., P. A., Cormick, Quest, D. O., Murtry, J. G., Riedel, C. J., Lazar, R. M., Stern, Y., & McConnolly, S. E. (2000). Severe pain confounds neuropsychological test performance. *Journal of Clinical and Experimental Neuropsychology*, *22*, 633–639.

Hicks, G. E., Gaines, J. M., Shardell, M., & Simonsick, E. M. (2008). Associations of back and leg pain with health status and functional capacity of older adults: Findings from the retirement community back pain study. *Arthritis Care & Research*, *59*, 1306–1313. https://doi.org/10.1002/art.24006

Hill, E. L., & Bird, C. M. (2006). Executive processes in asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia*, *44*, 2822–2835. https://doi.org/10.1016/j.neuropsychologia.2006.06.007

Holding, B. C., Ingre, M., Petrovic, P., Sundelin, T., & Axelsson, J. (2021). Quantifying cognitive impairment after sleep deprivation at different times of day: A proof of concept using ultra-short smartphone-based tests. *Frontiers in Behavioral Neuroscience*, *15*, 666146. https://doi.org/10.3389/fnbeh.2021.666146

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The real-world approach and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 721.

Jelicic, M., Henquet, C. E. C., Derix, M. M. A., & Jolles, J. (2001). Test-retest reliability of the behavioural assessment of the dysexecutive syndrome in a sample of psychiatric patients. *International Journal of Neuroscience*, *110*, 73–78. https://doi.org/10.3109/00207450108994222

Johnson, J. K., Lui, L., & Yaffe, K. (2007). Executive function, more than global cognition, predicts functional decline and mortality in elderly women. *The Journal of Gerontology: Series A*, *62*, 1134–1141.

Josman, N., Klinger, E., & Kizony, R. Construct Validity of the Virtual Action Planning-SupermarketVAP-S Comparison Between Healthy Controls and 3 Clinical Populations. In: *2009 Virtual Rehabilitation International Conference*, IEEE, 2009, 208–208. https://doi.org/10.1109/ICVR.2009.5174246

Jovanovski, D., Zakzanis, K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The multitasking in the city test. *Applied Neuropsychology: Adult*, *19*, 171–182. https://doi.org/10.1080/09084282.2011.643955

Jurica, P. J., Leitten, C. L., & Mattis, S. (2001). *Dementia Rating Scale-2^{TM} Professional Manual*. Psychological Assessment Resources.

Karzmark, P., Llanes, S., Tan, S., Deutsch, G., & Zeifert, P. (2012). Comparison of the frontal systems behavior scale and neuropsychological tests of executive functioning in predicting instrumental activities of daily living. *Applied Neuropsychology: Adult*, *19*, 81–85. https://doi.org/10.1080/09084282.2011.643942

Kenworthy, L., Freeman, A., Ratto, A., Dudley, K., Powell, K. K., Pugliese, C. E., Strang, J. F., Verbalis, A., & Anthony, L. G. (2020). Preliminary psychometrics for the executive function challenge task: A novel, "hot" flexibility, and planning task for youth. *Journal of the International Neuropsychological Society*, *26*, 725–732. https://doi.org/10.1017/S135561772000017X

Kievit, R. A., Romeijn, J.-W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Mind the Gap: A psychometric approach to the reduction problem. *Psychological Inquiry*, *22*, 67–87.

Kihlstrom, J. F. (2021). Ecological validity and "Ecological validity. *Perspectives on Psychological Science*, *16*, 466–471. https://doi.org/10.1177/1745691620966791

Killgore, W. D. S., Grugle, N. L., Reichardt, R. M., Killgore, D. B., & Balkin, T. J. (2009). Executive functions and the ability to sustain vigilance during sleep loss. *Aviation, Space, and Environmental Medicine*, *80*, 81–87. https://doi.org/10.3357/ASEM.2396.2009

Kraybill, M. L., & Suchy, Y. (2011). Executive functioning, motor programming, and functional independence: Accounting for variance, people, and time. *The Clinical Neuropsychologist*, *25*(2), 210–223. https://doi.org/10.1080/13854046.2010.542489.

Kraybill, M. L., Thorgusen, S. R., & Suchy, Y. (2013). The push-turn-taptap task outperforms measures of executive functioning in predicting declines in functionality: Evidence-Based approach to test validation. *The Clinical Neuropsychologist*, *27*, 238–255. https://doi.org/10.1080/13854046.2012.735702

Lamberts, K. F., Evans, J. J., & Spikman, J. M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology*, *32*, 56–65. https://doi.org/10.1080/13803390902806550

Lawton, M. P., & Brody, E. (1969). Assessment of older people: Self maintaining and instrumental activities of daily living. *The Gerontologist*, *9*, 179–186.

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (fifth ed.).Oxford University Press.

Liu, K. C. M., Chan, R. C. K., Chan, K. K. S., Tang, J. Y. M., Chiu, C. P. Y., Lam, M. M. L., Chan, S. K. W., Wong, G. H. Y., Hui, C. L. M., & Chen, E. Y. H. (2011). Executive function in first-episode schizophrenia: A three-year longitudinal study of an ecologically valid test. *Schizophrenia Research*, *126*, 87–92.

Long, C. J. (1996). Neuropsychological tests: A look at our past and the impact that ecological issues may have on our future. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 1–14). Gr Press/St Lucie Press, Inc, APA PsycInfo (1996-98718-001), https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-001&site=ehost-live, Retrieved from

Longaud-Valès, A., Chevignard, M., Dufour, C., Grill, J., Puget, S., Sainte-Rose, C., Valteau-Couanet, D., Dellatolas, G. (2016). Assessment of executive functioning in children and young adults treated for frontal lobe tumours using ecologically valid tests. *Neuropsychological Rehabilitation*, *26*, 558–583. https://doi.org/10.1080/09602011.2015.1048253

Mariani, E., Monastero, R., Ercolani, S., Rinaldi, P., Mangialasche, F., Costanzi, E., Vitale, D. F., Senin, U., & Mecocci, P. (2008). Influence of comorbidity and cognitive status on instrumental activities of daily living in amnestic mild cognitive impairment: results from the ReGA1 project. *International Journal of Geriatric Psychiatry*, *23*, 523–530. https://doi.org/10.1002/gps.1932

Miyata, S., Noda, A., Iwamoto, K., Kawano, N., Okuda, M., & Ozaki, N. (2013). Poor sleep quality impairs cognitive performance in older adults. *Journal of Sleep Research*, *22*, 535–541. https://doi.org/10.1111/jsr.12054

Moriyama, Y., Mimura, M., Kato, M., Yoshino, A., Hara, T., Kashima, H., Kato, A., Watanabe, A. (2002). Executive dysfunction and clinical outcome in chronic alcoholics. *Alcoholism, Clinical and Experimental Research*, *26*, 1239–1244. https://doi.org/10.1097/01.ALC.0000026103.08053.86

Ng, T. -P., Niti, M., Chiam, P. -C., & Kua, E. -H. (2006). Physical and cognitive domains of the instrumental activities of daily living: Validation in a multiethnic population of asian older adults. *The Journals of Gerontology: Series A: Biological Sciences and Medical Sciences*, *61*, 726–735. https://doi.org/10.1093/gerona/61.7.726

Nguyen, C. M., Copeland, C. T., Lowe, D. A., Heyanka, D. J., & Linck, J. F. (2020). Contribution of executive functioning to instrumental activities of daily living in older adults. *Applied Neuropsychology: Adult*, *27*, 326–333. https://doi.org/10.1080/23279095.2018.1550408

Niermeyer, M. A., & Suchy, Y. (2020). The vulnerability of executive functioning: The additive effects of recent non-restorative sleep, pain interference, and use of expressive suppression on test performance. *The Clinical Neuropsychologist*, *34*, 700–719. https://doi.org/10.1080/13854046.2019.1696892

Norris, G., & Tate, R. L. (2000). The behavioural assessment of the dysexecutive syndrome (BADS): Ecological, concurrent and construct validity. *Neuropsychological Rehabilitation*, *10*, 33–45.

O'Shea, R., Poz, R., Michael, A., Berrios, G. E., Evans, J. J., & Rubinsztein, J. S. (2010). Ecologically valid cognitive tests and everyday functioning in euthymic bipolar disorder patients. *Journal of Affective Disorders*, *125*, 336–340. https://doi.org/10.1016/j.jad.2009.12.012

Overdorp, E. J., Kessels, R. P. C., Claassen, J. A., & Oosterman, J. M. (2016). The combined effect of neuropsychological and neuropathological deficits on instrumental activities of daily living in older adults: A systematic review. *Neuropsychology Review*, *26*, 92–106. https://doi.org/10.1007/s11065-015-9312-y

Ownsworth, T., & McKenna, K. (2004). Investigation of factors related to employment outcome following traumatic brain injury: A critical review and conceptual model. *Disability and Rehabilitation: An International, Multidisciplinary Journal*, *26*, 765–784. https://doi.org/10.1080/09638280410001696700

Owsley, C., Sloane, M., McGwin Jr., G., & Ball, K. (2002). Timed instrumental activities of daily living tasks: Relationship to cognitive function and everyday performance assessments in older adults. *Gerontology*, *48*, 254–265. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12053117.

Perfetti, B., Varanese, S., Mercuri, P., Mancino, E., Saggino, A., & Onofrj, M. (2010). Behavioural assessment of dysexecutive syndrome in parkinson's disease without dementia: A comparison with other clinical executive tasks. *Parkinsonism & Related Disorders*, *16*, 46–50. https://doi.org/10.1016/j.parkreldis.2009.07.011APA PsycInfo (2009-14777-001).

Perna, R., Loughan, A. R., & Talka, K. (2012). Executive functioning and adaptive living skills after acquired brain injury. *Applied Neuropsychology: Adult*, *19*, 263–271. https://doi.org/10.1080/09084282.2012.670147

Putcha, D., & Tremont, G. (2016). Predictors of independence in instrumental activities of daily living: Amnestic versus nonamnestic MCI. *Journal of Clinical and Experimental Neuropsychology*, *38*, 991–1004. https://doi.org/10.1080/13803395.2016.1181716

Rabin, L. A., Burton, L. A., & Barr, W. B. (2007). Utilization rates of ecologically oriented instruments among clinical neuropsychologists. *The Clinical Neuropsychologist*, *21*, 727–743. https://doi.org/10.1080/13854040600888776

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: The virtual library task. *Journal of the International Neuropsychological Society*, *18*, 440–450. https://doi.org/10.1017/s1355617711001883

Robertson, K., & Schmitter-Edgecombe, M. (2017). Naturalistic tasks performed in realistic environments: A review with implications for neuropsychological assessment. *The Clinical Neuropsychologist*, *31*, 16–42. https://doi.org/10.1080/13854046.2016.1208847

Rochat, L., Ammann, J., Mayer, E., Annoni, J.-M., & Van der Linden, M. (2009). Executive disorders and perceived socio-emotional changes after traumatic brain injury. *Journal of Neuropsychology*, *3*, 213–227. https://doi.org/10.1348/174866408X397656

Romundstad, B., Solem, S., Brandt, A. E., Hypher, R. E., Risnes, K., Rø, T. B., Stubberud, J., & Finnanger, T. G. (2022). Validity of the behavioural assessment of the dysexecutive syndrome for children (BADS-C) in children and adolescents with pediatric acquired brain injury. *Neuropsychological Rehabilitation*, *33*, 551–573. https://doi.org/10.1080/09602011.2022.2034649

Rosenblum, S., Frisch, C., Deutsh-Castel, T., & Josman, N. (2015). Daily functioning profile of children with attention deficit hyperactive disorder: A pilot study using an ecological assessment. *Neuropsychological Rehabilitation*, *25*, 402–418. https://doi.org/10.1080/09602011.2014.940980

Rosetti, M. F., Ulloa, R. E., Reyes-Zamorano, E., Palacios-Cruz, L., de la Peña, F., & Hudson, R. (2018). A novel experimental paradigm to evaluate children and adolescents diagnosed with attention-deficit/hyperactivity disorder: Comparison with two standard neuropsychological methods. *Journal of Clinical and Experimental Neuropsychology*, *40*, 576–585. https://doi.org/10.1080/13803395.2017.1393501

Roy, A., Allain, P., Roulin, J.-L., Fournet, N., & Le Gall, D. (2015). Ecological approach of executive functions using the behavioural assessment of the dysexecutive syndrome for children (BADS-C): Developmental and validity study. *Journal of Clinical and Experimental Neuropsychology*, *37*, 956–971.

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity* *of neuropsychological testing* (pp. 15–41). Gr Press/St Lucie Press, Inc. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1996-98718-002&site=ehost-live

Schaeffer, M. J., Weerawardhena, H., Becker, S., & Callahan, B. L. (2022). Capturing daily-life executive impairments in adults: Does the choice of neuropsychological tests matter? *Applied Neuropsychology: Adult*, 1–11. https://doi.org/10.1080/23279095.2022.2109970

Schmitter-Edgecombe, M., Cunningham, R., McAlister, C., Arrotta, K., & Weakley, A. (2021). The night out task and scoring application: An ill-structured, open-ended clinic-based test representing cognitive capacities used in everyday situations. *Archives of Clinical Neuropsychology*, *36*, 537–553. https://doi.org/10.1093/arclin/acaa080

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following fronal lobe damage in man. *Brain*, *114*, 727–741.

Shimoni, M., Engel-Yeger, B., & Tirosh, E. (2012). Executive dysfunctions among boys with attention deficit hyperactivity disorder (ADHD): Performance-based test and parents report. *Research in Developmental Disabilities*, *33*, 858–865. https://doi.org/10.1016/j.ridd.2011.12.014

Spitoni, G. F., Aragonaa, M., Bevacqua, S., Cotugno, A., & Antonucci, G. (2018). An ecological approach to the behavioral assessment of executive functions in anorexia nervosa. *Psychiatry Research*, *259*, 283–288. https://doi.org/10.1016/j.psychres.2017.10.029

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, *21*, 327–337. https://doi.org/10.1016/j.acn.2006.04.004

Stamm, T. A., Pieber, K., Crevenna, R., & Dorner, T. E. (2016). Impairment in the activities of daily living in older adults with and without osteoporosis, osteoarthritis and chronic back pain: A secondary analysis of population-based health survey data. *BMC Musculoskeletal Disorders*, *17*, 139. https://doi.org/10.1186/s12891-016-0994-y

Stuss, D. T., & Knight, R. T. (Eds.). (2002). *Principles of Frontal Lobe Function*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195134971.001.0001

Suchy, Y. (2015). *Executive functions: A comprehensive guide for clinical practice*. Oxford University Press.

Suchy, Y., & Brothers, S. L. (2022). Reliability and validity of composite scores from the timed subtests of the D-KEFS battery. *Psychological Assessment*, *34*, 483–495. https://doi.org/10.1037/pas0001081

Suchy, Y., Brothers, S. L., Mullen, C., & Niermeyer, M. A. (2020). Chronic versus recent expressive suppression burdens are differentially associated with cognitive performance among older adults. *Journal of Clinical and Experimental Neuropsychology*, *42*, 834–848.

Suchy, Y., Lipio Brothers, S., DesRuisseaux, L. A., Gereau, M. M., Davis, J. R., Chilton, R. L. C., & Schmitter-Edgecombe, M. (2022). Ecological validity reconsidered: The night out task versus the D-KEFS. *Journal of Clinical and Experimental Neuropsychology*, *44*, 562–579. https://doi.org/10.1080/13803395.2022.2142527

Suchy, Y., Mullen, C., Brothers, S. L., & Niermeyer, M. A. (2020). Interpreting executive and lower-order error scores on the timed subtests of the Delis-Kaplan Executive Function System (D-KEFS) battery: Error analysis across the adult lifespan. *Journal of Clinical and Experimental Neuropsychology*, *42*, 982–997.

Suchy, Y., Niermeyer, M. A., Franchow, E. I., & Ziemnik, R. E. (2019). Naturally occurring expressive suppression is associated with lapses in instrumental activities of daily living among community-dwelling older adults. *Journal of the International Neuropsychological Society*, *25*, 718–728. https://doi.org/10.1017/S1355617719000328

Sudo, F. K., Alves, G. S., Ericeira-Valente, L., Alves, C. E. O., Tiel, C., Moreira, D. M., Laks, J., & Engelhardt, E. (2015). Executive testing predicts functional loss in subjects with white matter lesions. *Neurocase*, *21*, 679–687. https://doi.org/10.1080/13554794.2014.973884

Tinajero, R., Williams, P. G., Cribbet, M. R., Rau, H. K., Bride, D. L., & Suchy, Y. (2018). Nonrestorative sleep in healthy, young adults without insomnia: Associations with executive functioning, fatigue, and pre-sleep arousal. *Sleep Health*, *4*, 284–291. https://doi.org/10.1016/j.sleh.2018.02.006

Torralva, T., Strejilevich, S., Gleichgerrcht, E., Roca, M., Martino, D., Cetkovich, M., & Manes, F. (2012). Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. *Bipolar Disorders*, *14*, 118–125. https://doi.org/10.1111/j.1399-5618.2012.00987.x

Valls-Serrano, C., Verdejo-García, A., Noël, X., & Caracuel, A. (2018). Development of a contextualized version of the multiple errands test for people with substance dependence. *Journal of the International Neuropsychological Society*, *24*, 347–359. https://doi.org/10.1017/S1355617717001023

Van Voorhis, W., Carmen, R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorial in Quantitative Methods for Psychology*, *3*, 43–50.

Verdejo-García, A., & Pérez-García, M. (2007). Ecological assessment of executive functions in substance dependent individuals. *Drug and Alcohol Dependence*, *90*, 48–55. https://doi.org/10.1016/j.drugalcdep.2007.02.010

Webb, C. A., Cui, R., Titus, C., Fiske, A., & Nadorff, M. R. (2018). Sleep disturbance, activities of daily living, and depressive symptoms among older adults. *Clinical Gerontologist*, *41*, 172–180. https://doi.org/10.1080/07317115.2017.1408733

Werner, P., Rabinowitz, S., Klinger, E., Korczyn, A. D., & Josman, N. (2009). Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment: A preliminary study. *Dementia and Geriatric Cognitive Disorders*, *27*, 301–309. https://doi.org/10.1159/000204915

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on "open-ended" executive function tests in autism. *Autism Research*, *2*, 138–147. https://doi.org/10.1002/aur.78

Williams, P. G., Suchy, Y., & Rau, H. K. (2009). Individual differences in executive functioning: Implications for stress regulation. *Annals of Behavioral Medicine*, *37*, 126–140. https://doi.org/10.1007/s12160-009-9100-0

Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied & Preventive Psychology*, *2*, 209–215. https://doi.org/10.1016/S0962-1849(05)80091-5

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with dysexecutive syndrome. *Neuropsychological Rehabilitation*, *8*, 213–228.

Yesavage, J. A. (1988). Geriatric Depression Scale. *Psychopharmacology Bulletin*, *24*, 709–711. National Institute of Mental Health.

Zartman, A. L., Hilsabeck, R. C., Guarnaccia, C. A., & Houtz, A. (2013). The pillbox test: An ecological measure of executive functioning and estimate of medication management abilities. *Archives of Clinical Neuropsychology*, *28*, 307–319. https://doi.org/10.1093/arclin/act014

Ziemnik, R. E., & Suchy, Y. (2019). Ecological validity of performance-based measures of executive functions: Is face validity necessary for prediction of daily functioning? *Psychological Assessment*, *31*, 1307–1318.https://doi.org/10.1037/pas0000751