

# Reliability and Validity of Cognitive Workload in Older Adults

Hannes Devos<sup>1\*</sup>, Kathleen Gustafson<sup>2,3</sup>, Pedram Ahmadnezhad,<sup>1</sup> Ke Liao<sup>3</sup>, Jonathan D. Mahnken,<sup>4,5</sup> William M. Brooks,<sup>2,3,4</sup> Jeffrey M. Burns<sup>2,4</sup>

1 <sup>1</sup>Laboratory for Advanced Rehabilitation Research in Simulation, Department of Physical Therapy  
2 and Rehabilitation Science, University of Kansas Medical Center, Kansas City, KS, USA

3 <sup>2</sup>Department of Neurology, University of Kansas Medical Center, Kansas City, KS

4 <sup>3</sup>Hoglund Brain Imaging Center, University of Kansas Medical Center, Kansas City KS

5 <sup>4</sup>University of Kansas Alzheimer's Disease Center, University of Kansas Medical Center, Kansas  
6 City, KS

7 <sup>5</sup>Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas City, KS

8

9 \* **Correspondence:**

10 Hannes Devos

11 Department of Physical Therapy and Rehabilitation Science, University of Kansas Medical Center,  
12 MS2002, 3901 Rainbow Blvd., Kansas City KS

13 E-mail: hdevos@kumc.edu

14 **Keywords: event-related potentials, pupils, workload, reliability, working memory, mild**  
15 **cognitive impairment, dementia, preclinical (Min.5-Max. 8)**

16 **Abstract**

17 Cognitive workload (mental effort) is a measure of attention allocation to a task, which can be  
18 administered through self-report or physiological measures. Cognitive workload is increasingly  
19 recognized as an important determinant of performance in cognitive tests and daily life activities.  
20 However, the reliability and validity of these measures have not been established in older adults with  
21 a wide range of cognitive ability. The aim of this study was to establish the test-retest reliability of  
22 the NASA-Task Load Index (NASA-TLX) and Index of Cognitive Activity (ICA), extracted from  
23 pupillary size. The convergent validity of these measures against event-related potentials (ERPs) was  
24 also investigated. A total of 38 individuals with scores on the Montreal Cognitive Assessment  
25 ranging between 17 and 30 completed a working memory test (n-back) with three levels of difficulty  
26 at baseline and two-week follow-up. Intraclass correlation coefficients (ICC) values of the NASA-  
27 TLX ranged between 0.71 and 0.81, demonstrating good to excellent reliability. Mean ICA scores  
28 showed fair to good reliability, with ICC's ranging between 0.56 and 0.73. Mean ICA and NASA-  
29 TLX scores showed significant and moderate correlations (Pearson r range between 0.30 and 0.33)  
30 with P3 ERP at the midline channels. We conclude that ICC and NASA-TLX are reliable measures  
31 of cognitive workload in older adults. Further research is needed in dissecting the subjective and  
32 objective constructs of cognitive workload.

33

## 34 **1 Introduction**

35 Despite its incredible power and flexibility, there are limits to the brain’s capabilities. For example,  
36 working memory—the storage space that provides the foundation for higher-order cognitive  
37 functions—is on average only able to retain four items at any given time.[1] Performance on working  
38 memory tests is determined by the brain’s ability to allocate attention (mental effort) to the task and  
39 its available resources. Mental effort has been conceptualized as “cognitive workload” and if  
40 measured well may have properties that offer relevant information beyond that provided by standard  
41 performance measures such as accuracy or response times of a task. Cognitive workload has  
42 traditionally been characterized as a direct measure of attention allocation to the task,[2] although  
43 other studies have postulated that cognitive workload better reflects the readiness for resource  
44 expenditure.[3] When the cognitive workload required by the task is lower than the available  
45 cognitive resources, the task has the potential to be executed successfully. When the cognitive  
46 workload imposed by the task exceeds the available resources, task performance is expected to  
47 decrease.[3] Older age and age-related neurodegeneration may affect the availability of cognitive  
48 resources. With fewer resources available to attend to the task, older adults may show greater  
49 workload on a task compared to younger individuals.[3] This increased cognitive workload may  
50 reflect inefficient or compensatory use of neural resources to cope with the demand of the task. Some  
51 studies have suggested that this increased cognitive workload may serve as a predictor of cognitive  
52 decline.[4]

53 Several techniques have been developed to measure cognitive workload, including questionnaires,  
54 performance outcomes, or physiological measures. The National Aeronautics and Space  
55 Administration – Task Load Index (NASA-TLX)[5] is one of the most widely used questionnaires of  
56 cognitive workload.[6] This questionnaire relies on self-recall of cognitive workload and is typically  
57 administered after completion of the task. The NASA-TLX therefore does not provide continuous  
58 data but relies on participant’s memory and self-recall of events that have already occurred. Although  
59 the psychometric properties of the NASA-TLX have been established in a variety of disciplines such  
60 as aviation, military, driving, or skill acquisition, [7 8] the reliability and validity of this instrument  
61 have not been tested in older adults with different levels of cognitive functioning.

62 Performance measures such as accuracy and response times are considered indirect measures of  
63 cognitive workload expenditure because they do not directly capture brain activity. A previous study  
64 by our group found accuracy and response times on the n-back test to be highly reliable performance  
65 measures of working memory in older adults.[9] Unlike performance measures, physiological  
66 measures can provide a continuous recording of brain activity in real time. Some studies have  
67 suggested that physiological changes may appear before manifestation of symptoms in performance  
68 measures, thus providing a more sensitive measure of early cognitive decline.[10-12] In a systematic  
69 review, Ranchet et al (2017) scrutinized the physiological changes resulting from increased cognitive  
70 workload in older adults with and without cognitive impairment. Increased hemodynamic and  
71 electrophysiological activity in the brain, smaller changes in systolic blood pressure, and increased  
72 pupillary dilation were observed in healthy older adults compared to younger adults, suggesting  
73 additional recruitment of neural resources to cope with task demand. In adults with  
74 neurodegenerative conditions, the inability to cope with task demand was even more apparent,  
75 resulting in not only an increase in hemodynamic, electrophysiological, and pupillary responses, but  
76 also worsening on performance measures.[4]

77 Of those, the pupillary response is particularly interesting since it has been implicated with early tau  
78 accumulation in the locus coeruleus (LC) in Alzheimer's disease (AD).[13] Decreased neuronal  
79 density of the LC has been associated with cognitive decline in older adults, mild cognitive  
80 impairment, and AD.[14] The LC plays an essential role in the regulation of physiological arousal  
81 and cognition.[15] When activated, the LC sends inhibitory projections to the parasympathetic  
82 Edinger-Westphal nucleus, which in turn inhibits contraction of the pupillary sphincter muscle.[16]  
83 LC activity also triggers the sympathetic nervous system, resulting in activation of the pupillary  
84 dilator muscle.[17] A previous study found increased pupillary dilation in participants with single-  
85 domain mild cognitive impairment compared to cognitively normal participants, despite performance  
86 in the normal ranges.[12] Furthermore, participants with a genetic predisposition for AD were more  
87 likely to exhibit larger pupillary size in highly cognitive demanding tasks.[18]

88 Two methods of pupillary response to cognitive workload have been reported. Task-evoked pupillary  
89 response (TEPR) compares the averaged raw pupillary diameter after stimulus onset to the averaged  
90 baseline pupillary diameter. Using raw pupillary dilation as a measure of cognitive workload poses  
91 some challenges as the light reflex may confound extracting the TEPR, especially in settings where  
92 the lighting of the surrounding environment or the luminosity of the screen cannot be entirely  
93 controlled.[19] Changes in camera angle and eye movements may also interfere with raw pupillary  
94 recording.[17 20] Nonetheless, previous studies found increased TEPR in individuals with elevated  
95 risk of AD.[18 21] An alternative to pupillometric baseline-related difference measures is the  
96 moment-to-moment pupillary diameter measurement. The Index of Cognitive Activity (ICA) and the  
97 Index of Pupillary Activity (IPA) are two moment-to-moment measures that calculate the rate of  
98 change of pupillary diameter rather than the difference between averaged pupillary diameter after and  
99 before stimulus onset.[22 23] Both ICA and IPA measures are based on the premise that pupils  
100 continuously undergo small fluctuations, even in steady illumination conditions.[24] An increase in  
101 abrupt discontinuities in the small oscillatory movements of the pupil reflects increased cognitive  
102 workload. These two measures of cognitive workload are claimed to successfully separate the  
103 pupillary response to cognitive workload from the light reflex. Furthermore, the ICA claims to be  
104 unaffected by changes in eye movements and sampling rate.[25] The ICA in particular has been used  
105 to investigate changes in cognitive workload in individuals at risk of cognitive impairment, including  
106 Parkinson's disease, multiple sclerosis, and breast cancer.[4 26-30] Overall, the ICA seems to  
107 increase with cognitive demand, regardless of disease condition[27]. In addition, some studies report  
108 that individuals with increased risk of cognitive impairment show greater ICA compared to  
109 controls.[28 30 31] However, the reliability and validity of ICA during working memory tasks in  
110 older individuals have not been established.

111 There is no gold standard for measuring cognitive workload. We selected the P3 (or P300) event-  
112 related potential (ERP) as our criterion measure. The P3 is a positive peak at around 300 ms observed  
113 in visual or auditory oddball tasks of working memory. This component is considered a sensitive and  
114 reliable measure of cognitive workload, including in older adults with cognitive impairment.[4 9 32-  
115 34] In addition, the P3 ERP is assumed to share the same neural origins in the LC as the pupillary  
116 response to cognitive workload, making this physiological response particularly suitable as a  
117 criterion measurement.[35 36]

118 The aim of this study was to demonstrate the reliability and convergent validity of the NASA-TLX  
119 and the ICA in older adults with a wide range of cognitive ability.

## 120 **2 Materials and Methods**

### 121 **2.1 Participants**

122 In this test-retest reliability study, 38 right-handed participants were recruited from the University of  
123 Kansas Alzheimer's Disease Center between 05/03/2018 and 03/10/2020. Participants were included  
124 in the study if they (1) signed informed consent; (2) were 65 years of age or older; and (3) able to  
125 understand the instructions in English. Exclusion criteria were: (1) currently taking steroids,  
126 benzodiazepines, or neuroleptics; (2) history of any substance abuse, (3) history of a psychiatric or  
127 neurological disorder other than MCI or AD; and (4) vision problems that cannot be resolved by  
128 corrective lenses.

129 Each participant had previously undergone an amyloid PET scan of the brain. Intravenous florbetapir  
130 F-18A was administered in a GE Discovery ST-16 PET/CT scanner to assess cerebral amyloid  
131 burden. Standard Uptake Value Ratio for six regions of interest was calculated using MIMneuro  
132 software (MiM Software Inc, Cleveland, OH) by normalizing the A $\beta$  PET image to the entire  
133 cerebellum. Participants were categorized in one of three groups: (1) non-elevated or A $\beta$ -; (2)  
134 elevated or A $\beta$ +; or (3) MCI / AD. The recommendations from NIA and the Alzheimer's Association  
135 workgroup were used to categorize participants into A $\beta$ - and A $\beta$ +. [37] The protocol for  
136 determination of amyloid elevation is described elsewhere. [38] The average (standard deviation) time  
137 between administration of PET scan and pupillometry/EEG assessment was 1090 (479) days. Sixteen  
138 were cognitively normal older adults with no elevated amyloid PET scans (A $\beta$ -), 16 were cognitive  
139 normal with elevated amyloid PET scans (A $\beta$ +), and six had a clinical diagnosis of MCI or AD.  
140 Participants completed their two-week follow-up session  $16 \pm 7$  days after the first session. Each  
141 session lasted about 60 minutes including rest breaks.

## 142 **2.2 Procedure**

### 143 **2.2.1 Demographic and Clinical Information**

144 Age, sex, and education were recorded. General cognitive functions were evaluated with the  
145 Montreal Cognitive Assessment (MOCA). [39] Scores on the MOCA ranged between 0 and 30.

### 146 **2.2.2 N-back Test**

147 In this study, the 0-back, 1-back, and 2-back tests were administered. The 0-back test is essentially a  
148 memory search task of sustained attention and often used as a control condition. [40 41]. Participants  
149 were instructed to press the button as soon as the letter "X" amongst a series of distracter letters  
150 appeared on the screen. The 1-back test requires the participant to passively store and update  
151 information in working memory. In this test, participants had to press the button if the current letter  
152 was the same as the previous letter. The 2-back test requires continuous mental effort to update  
153 information of new stimuli and maintain representations of recently presented stimuli in short-term  
154 memory. [42] Participants were instructed to press the button when the current letter was the same as  
155 the letter presented 2 places before.

156 Extensive description of the 7-minutes test is provided elsewhere. [9] In short, each n-back test  
157 comprised 180 trials, including 60 (33.3%) target trials and 120 (66.7%) nontarget trials. Display  
158 time of each letter was 500 ms, followed by a blank interstimulus interval of 1700 ms with a random  
159 jitter of 50 ms. Maximum response time was 2150 ms. The participants practiced before the task.

### 160 **2.2.3 National Aeronautics and Space Administration Task Load Index**

161 The NASA-TLX is one of the most used self-reported questionnaires of cognitive workload. Six  
162 items of mental demand, physical demand, temporal demand, effort, performance, and frustration  
163 provide a comprehensive measure of cognitive workload. [43] Each item is scored on a visual

164 analogue scale ranging from 0 to 100 in 5-point increments. NASA-TLX was administered  
165 immediately after each n-back test. The mean score of the six subscales was computed for each of the  
166 conditions and for each subject. In contrast to the original calculation,[5] we did not attribute weights  
167 to each of the components since the unweighted average produced better sensitivity and reliability  
168 than the weighted average.[44]

#### 169 **2.2.4 Index of Cognitive Activity**

170 While doing the n-back test, participants wore mobile eye tracking glasses (SMI ETG 2,  
171 Sensomotoric Instruments, Teltow, Germany). Pupillary size was recorded in real-time at 60 Hz  
172 using infrared cameras for both the left and right eye. Pupillary data were analyzed using Eyeworks  
173 (Eye Tracking, Inc, Solana Beach, CA, USA). The software analyzed the change in pupil size for  
174 each eye throughout each n-back test. Potential artifacts from lighting and eye movements were  
175 minimized by using constant room lighting and having the participants focus on the screen. However,  
176 even under constant lighting conditions, the pupil continues to oscillate irregularly. Therefore, we  
177 transformed raw pupil data to Index of Cognitive Activity (ICA) scores. The ICA discriminates rapid,  
178 small bursts in pupillary dilation due to cognitive workload from slower, larger amplitude changes in  
179 pupillary size due to the light reflex by decomposing the raw pupillary size to different wavelets of  
180 high and low frequency components of the signal. The ICA has a low autocorrelation at a lag of 100  
181 ms, and almost no autocorrelation at a lag of 200 ms. The short latency features and low  
182 autocorrelation make the ICA particularly suitable for oddball tasks.[45] The ICA is calculated by  
183 dividing the number of rapid small pupillary dilations per second by the number of expected rapid  
184 pupillary dilations per second. The values are then transformed using the hyperbolic tangent function.  
185 Blinks are factored out by linear interpolation of adjacent time spans to produce continuous values  
186 ranging between 0 and 1.  
187 The average percentage of missing data collected from the eye tracker ranged between 0.87% and  
188 2.24%. Three participants had more than 50% of missing ICA values in one or more tests. These  
189 values were excluded from the analyses. Mean ICA of the left and right eye were included as  
190 outcome measures.

#### 191 **2.2.5 P3 Event-Related Potential**

192 Continuous electro-encephalogram (EEG) was recorded at 1,000 Hz using an Electrical Geodesics  
193 high-density system (Magstim EGI, Eugene, OR, USA) with 256 scalp electrodes. The start and end  
194 of the task were time-stamped and synchronized with EEG and ICA recordings. EEG recordings  
195 were filtered from 0.50 to 30 Hz using EGI software. All other EEG processing was done in  
196 EEGLab[46] and in ERPLab.[47] EEG data were online referenced to Cz and offline re-referenced to  
197 average of mastoids. Cz was interpolated using the surrounding five channels. Independent  
198 component analysis was employed to separate brain activity from ocular, muscular, or cardiovascular  
199 artifacts. Signals from bad electrodes were removed and interpolated with the data of surrounding  
200 electrodes. Continuous EEG data were segmented into epochs ranging between -100 and 1000 ms of  
201 stimulus onset. Each epoch was baseline corrected using the prestimulus interval. Scalp locations and  
202 measurement windows for the P3 ERP were based on their spatial extent and latency after inspection  
203 of grand average waveform of the task effect. The task effect was calculated by subtracting the  
204 average ERP elicited from the targets from the average ERP elicited by non-targets for each  
205 participant. The P3 component time window was established between 200 ms and 400 ms for all  
206 three tests. Because of the prefrontal cortex involvement in working memory, we identified *a priori*  
207 Fz as the main channel, but also report results of the midline electrodes Cz and Pz. No participants  
208 were removed from the analyses because of artifacts. P3 peak amplitude of the task effect was

## Psychometric Properties of Cognitive Workload in Older Adults

209 considered the main outcome measure to test convergent validity against, but we also calculated P3  
 210 peak latency. P3 peak amplitude, and to a lesser extent P3 peak latency, are reliable measures of  
 211 cognitive workload.[9]

### 212 Data Analysis

213 Descriptive analysis including mean (standard deviation, SD) and frequency count of participants'  
 214 general, performance measures, NASA-TLX, ICA, and ERP data were performed as appropriate.  
 215 Intra-class correlation coefficients (ICC) were used to calculate test-retest reliability of ICA values  
 216 and NASA-TLX scores. ICCs were computed as the between subject variance divided by the total  
 217 (between + within) variance.[48] ICC values less than 0.40 were considered poor; values between  
 218 0.40 and 0.59 fair, values between 0.60 and 0.74 good, and values between 0.75 and 1.00  
 219 excellent.[49]. Bland-Altman plots were used to visualize the measurement precision of ICA values  
 220 and NASA-TLX scores across the test moments.[50] Intersubject stability according to subject  
 221 rankings was calculated using the Pearson r correlation coefficient. Minimal Detectable Change at a  
 222 90% confidence interval (MDC<sub>90</sub>) provides an clinically useful indication of absolute reliability and  
 223 reflects whether an observed change score is above that expected due to measurement error.[51]  
 224 MDC<sub>90</sub> was calculated as 1.645 x standard error of measurement (SEM) x  $\sqrt{2}$  where SEM = SD<sub>(first</sub>  
 225 test) x  $\sqrt{(1 - ICC)}$ . The Kolmogorov–Smirnov test was employed to test the normality of our data  
 226 distribution in addition to visualization of Q-Q plots. All analyses were done using SAS Enterprise  
 227 8.2 and SAS 9.4 software. The threshold of significance was set at  $\alpha = 0.05$ .

## 228 3 Results

### 229 3.1 Participant Characteristics

230 Participants (n = 38; 23 (61%) women) were on average 73.81 (5.23) years old and scored 26.97  
 231 (2.91) the MOCA scale. MOCA scores of participants ranged between 17 and 30.

### 232 3.2 Test-Retest Reliability of NASA

233 Overall, the NASA-TLX scores showed great consistency across the two test moments. ICC scores  
 234 ranged between 0.71 for 2-back and 0.81 for 0-back, demonstrating good to excellent reliability  
 235 (Table 1). Pearson r correlations ranged between 0.55 for 2-back and 0.68 for 0-back, indicating  
 236 strong intersubject stability. MDC of the NASA-TLX ranged from 15.82 points on the 0-back test to  
 237 24.33 points on the 2-back.  
 238

Table 1. Comparison of NASA-TLX and ICA at baseline and two-week follow-up (n = 38).

Variable	Baseline	Follow-up	Pearson r	ICC, (95% CI)	MDC <sub>90</sub>
0-back, NASA-TLX	19.51 (15.95)	21.98 (17.89)	0.68 <sup>a</sup>	0.81 (0.61 – 0.90) <sup>a</sup>	15.82
0-back, mean ICA L	0.33 (0.14)	0.24 (0.15)	0.46 <sup>b</sup>	0.63 (0.26 – 0.83) <sup>b</sup>	0.20
0-back, mean ICA R	0.27 (0.17)	0.28 (0.16)	0.55 <sup>b</sup>	0.70 (0.40 – 0.85) <sup>a</sup>	0.22
1-back, NASA-TLX	28.24 (17.80)	27.22 (16.96)	0.60 <sup>a</sup>	0.78 (0.57 – 0.89) <sup>a</sup>	19.37
1-back, mean ICA L	0.29 (0.17)	0.27 (0.14)	0.58 <sup>a</sup>	0.73 (0.47 – 0.86) <sup>a</sup>	0.20
1-back, mean ICA R	0.25 (0.16)	0.30 (0.15)	0.39 <sup>c</sup>	0.56 (0.12 – 0.78) <sup>b</sup>	0.25

2-back, NASA-TLX	50.92 (19.41)	50.61 (19.13)	0.55 <sup>b</sup>	0.71 (0.42 – 0.85) <sup>a</sup>	24.33
2-back, mean ICA L	0.25 (0.14)	0.23 (0.16)	0.50 <sup>b</sup>	0.64 (0.29 – 0.82) <sup>a</sup>	0.24
2-back, mean ICA R	0.24 (0.18)	0.25 (0.15)	0.45 <sup>b</sup>	0.62 (0.20 – 0.82) <sup>b</sup>	0.25

Abbreviations: CI, confidence interval; ICA, Index of Cognitive Activity; ICC, Intraclass correlation coefficient; L, left, NASA-TLX, MDC, minimal detectable difference; NASA Task Load Index; R, right.

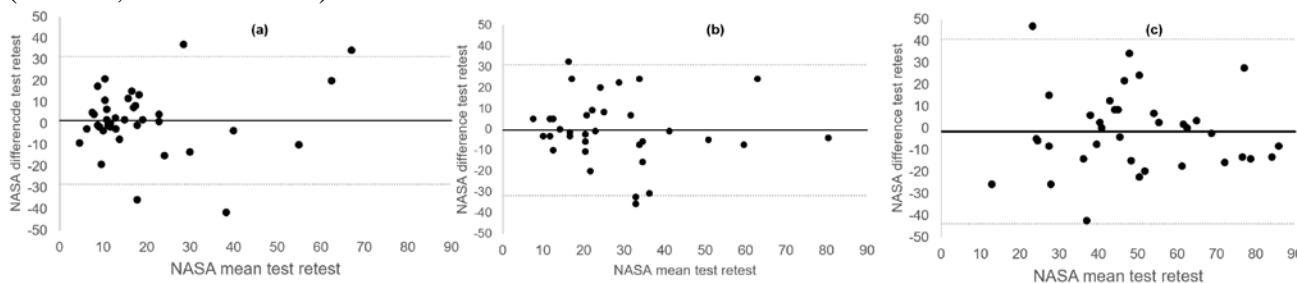
<sup>a</sup>p<0.0001

<sup>b</sup>p<0.01

<sup>c</sup>p<0.05

239

240 Bland-Altman plots showed equal spread of data around the mean (Figure 1). However, the spread of  
 241 NASA-TLX difference scores (limits of agreement, LOA) was slightly larger in the 2-back test (95%  
 242 confidence interval (CI), -43.88 – 40.63) compared to 0-back (95% CI, -28.37 – 31.28) and 1-back  
 243 (95% CI, -31.78 – 30.95) tests.



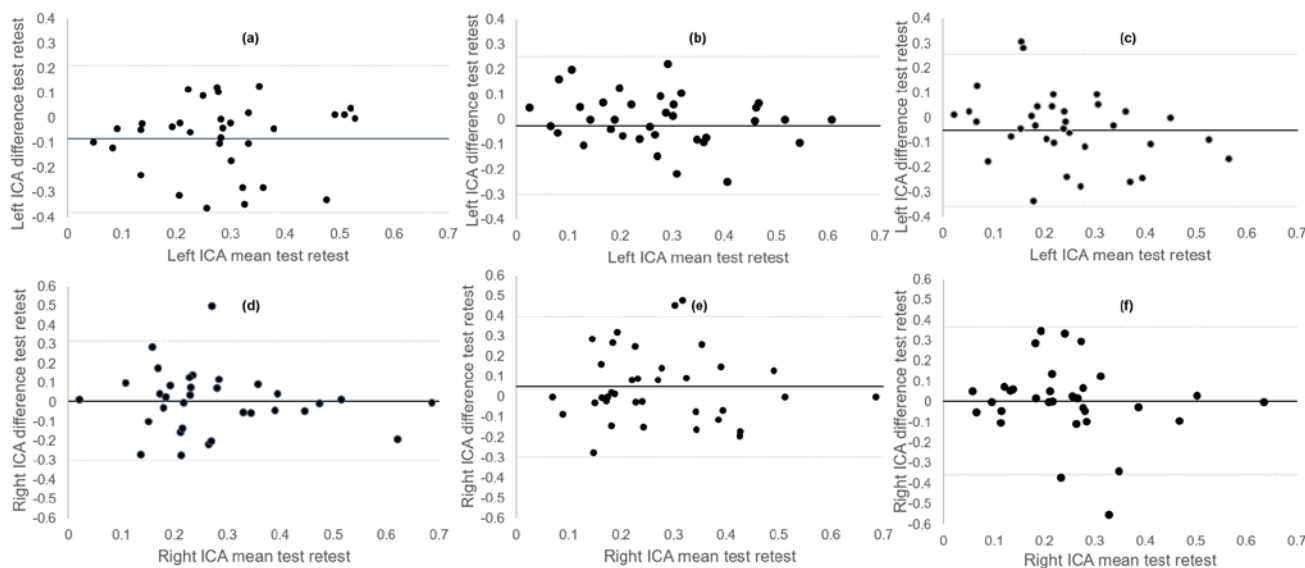
244

245 Figure 1. Bland Altman Plots of (a) 0-back NASA-TLX; (b) 1-back NASA-TLX; (c) 2-back NASA-  
 246 TLX.

### 247 3.3 Test-Retest Reliability of ICA

248 All ICC values of the ICA measure were statistically significant (Table 1). ICC values ranged from  
 249 0.46 for mean ICA right eye in the 1-back test to 0.73 for mean ICA left eye in the 1-back test. All  
 250 ICC values produced fair to good reliability. Pearson r correlations ranged from 0.39 (mean ICA  
 251 right eye in the 1-back test) to 0.58 (mean ICA left eye in 1-back). MDC of ICA values ranged from  
 252 0.20 (in 0-back and 1-back) to 0.25 in 2-back for the right eye.

253 Figure 2 shows the Bland-Altman plots for mean ICA in the left eye for each test. Plot (a) showed a  
 254 slight tendency towards practice effect in the mean ICA of the left eye during the 0-back test, with  
 255 decreased ICA at 2-week follow-up compared to baseline assessment. Plots (b) to (f) demonstrated  
 256 equal distribution of the data around zero, indicating no bias in the results and no heteroscedasticity  
 257 within the data.



258

259 Figure 2. Bland Altman Plots of (a) 0-back ICA mean left eye (b) 1-back ICA mean left eye; (c) 2-  
260 back ICA mean left eye; (d) 0-back ICA mean right eye; (e) 1-back ICA mean right eye; (f) 2-back  
261 ICA mean right eye. ICA, Index of Cognitive Activity.

### 262 3.4 Convergent validity of NASA-TLX

263 There was a trend that higher scores on the NASA-TLX correlated with increased peak P3 latency at  
264 channel Fz ( $r = 0.31$ ;  $p = 0.06$ ) for the 0-back.

265 Similar results were found for the 1-back test. Higher NASA-TLX scores correlated with increased  
266 peak P3 latency at Pz ( $r = 0.32$ ;  $p = 0.05$ ).

267 No correlations were found between NASA and ERP measures for the 2-back.

### 268 3.5 Convergent validity of ICA

269 ICA (mean of left and right eyes) were correlated to ERP (amplitude and latency of Fz, Cz, and Pz).

270 No significant correlations were found between ICA and ERP in 0-back.

271 Larger mean ICA in the right eye correlated significantly with increased P3 peak latency at Fz ( $r =$   
272  $0.32$ ;  $p = 0.049$ ) and at Cz ( $r = 0.33$ ;  $p = 0.048$ ) in the 1-back test. Likewise, larger mean ICA in the  
273 left eye correlated with increased P3 peak latency at Pz ( $r = 0.35$ ;  $p = 0.03$ ), and with larger P3 peak  
274 amplitude at Cz ( $r = 0.32$ ;  $p = 0.048$ ).

275 Larger mean ICA of the left eye correlated with increased P3 peak latency in Pz ( $r = -0.34$ ;  $p = 0.04$ )  
276 in the 2-back.

## 277 4 Discussion

278 Our results showed that pupillary response, transformed to an Index of Cognitive Activity (ICA),  
279 provides fair to good test-retest reliability as a measure of real-time cognitive workload in older  
280 adults with and without cognitive impairments. Subjective measures, such as the NASA-TLX, offers



281 even better reliability of cognitive workload in older adults. Moderate correlations were found  
282 between these two measures and the P3 ERP.

283  
284 The ease of use of the NASA-TLX has resulted in applications in diverse fields of aviation, military,  
285 human-machine interaction, driving, and medicine.[7] Despite the vast literature, few studies have  
286 reported on the test-retest reliability of the NASA-TLX in healthy adults and none in older adults  
287 with or without cognitive impairments. Battiste and Bortolussi reported strong test-retest reliability ( $r$   
288 = 0.77) of the NASA-TLX in airborne pilots. Hart and Staveland found a correlation of 0.83 in  
289 NASA scores administered at baseline and four week follow-up assessment in healthy adults.[5] Xiao  
290 reported a test-retest reliability of 0.75 in mental health workers.[52] These correlations coefficients  
291 are slightly higher than those found in our study (ranging between 0.55 and 0.68), which may be  
292 because our study focused on older adults with a wide range of cognitive ability. Previous studies  
293 have shown a potential confounding effect of cognitive impairment on reliability of EEG ERP.[9 53]  
294 However, Pearson correlation coefficients tend to overestimate the true test-retest reliability. We  
295 extended the correlation analyses with intra-class correlation coefficients (ICC), Bland-Altman plots,  
296 and minimal detectable change (MDC) calculations. ICC values provide a single measure of the  
297 magnitude of agreement while accounting for the differences in test moments along with the  
298 correlation between test moments. The ICC's showed good to excellent reliability for the NASA-  
299 TLX and no signs of systematic bias across the two test moments. The ICC's were higher and the  
300 range scores were smaller (0.71 – 0.81) than those reported in a previous study (range 0.34 – 0.80)  
301 that taxed the mental and physical effort of simulated manufacturing tasks in 24 college engineering  
302 students.[44]. None of the aforementioned studies provided a graphical representation of the  
303 measurement error across the two test moments. The Bland-Altman graphs showed relatively large  
304 limits of agreement, with no evidence of test or practice effect in all three tests. MDC calculations  
305 showed changes of 15% to 25% of total NASA-TLX scale scores to represent true change beyond  
306 measurement error. Taken together, subjective self-recall of cognitive workload is reliable across the  
307 spectrum of cognitive aging and has potential to be used as a measure of mental effort expenditure in  
308 this population.

309  
310 Likewise, pupillometry has been used for over five decades as a measure of cognitive workload in the  
311 domains of psychophysiology, cognitive neuroscience, and human factors engineering such as  
312 aviation or driving. Only recently has pupillometry received attention in the medical field as a  
313 potential marker of disease progression in adults with Alzheimer's disease, Parkinson's disease, and  
314 breast cancer.[12 18 27 30] This rekindled interest in pupillary response to cognitive workload as a  
315 marker of cognitive decline warrants an investigation of its psychometric properties. Overall, the ICA  
316 produced fair to good reliability scores, ranging between 0.56 and 0.78. These ICC's are in the same  
317 range as the reliability of our convergent measure, the P3 ERP component.[9] Comparison of  
318 reliability with other measures of pupillary response is complicated by the type of extraction (TEPR  
319 versus ICA), the type of task, and the population of interest. The closest comparison is the study by  
320 Kahya et al, that estimated reliability of ICA during postural demanding tasks in Parkinson's  
321 disease.[54] The ICC's in that study ranged between 0.74 and 0.93, which are higher than those  
322 reported in the current study. However, participants in that study completed the retest within hours of  
323 the first test, which may have resulted in less day-to-day variability. The Bland-Altman plots  
324 revealed no systematic bias of ICA across test moments. MDC values showed a change between 20%  
325 and 25% of the total scale score is needed to produce an effect that cannot be attributed to  
326 measurement error. These results suggest that ICA provides a stable measure of cognitive workload  
327 during cognitive testing in older adults with and without cognitive impairments.

328

329 NASA-TLX and ICA correlated only moderately with P3 ERP. A previous study demonstrated a  
330 strong correlation ( $r = -0.70$ ) between peak ICA values and ERP P3 latency in healthy young  
331 adults.[55] Comparison of our results with this study is complicated since different ICA metrics  
332 (mean versus peak), ERP measures (amplitude versus peak), and population (older versus younger)  
333 were used. In addition, this study used a measure of working memory whereas the other study used a  
334 cognitive-motor interference balance task. Although n-back is arguably the most ubiquitous working  
335 memory test used in ERP studies across the age spectrum,[40] previous studies have shown that the  
336 n-back test hosts an array of control processes, including speed of processing, storage, comparison  
337 processes, updating, keeping track, task mixing, task shifting, and resistance to interference.[40 41  
338 56] Therefore, the n-back test is a multi-domain cognitive assessment rather than a single-domain test  
339 of working memory. These multidomain processes involved with the n-back test may explain the  
340 moderate correlations between the ICA and P3 ERP. An alternative explanation is that cognitive  
341 workload represents several dimensions of mental, physical, and temporal demand, along with effort,  
342 performance, and frustration. It may be that ICA and P3 ERP measure overlapping, yet distinct  
343 constructs of cognitive workload. This assumption should be tested in future studies. It also remains  
344 unclear why left and right ICA values produced different correlations. While some studies have  
345 suggested a lateralization effect of hemispheric function on pupillary response,[57] in this case, the  
346 differences are likely due to measurement error.

347  
348 To our knowledge, this is the first study evaluating the psychometric properties of a subjective and  
349 objective measure of cognitive workload in group of older adults with a heterogeneous profile of  
350 cognitive ability. We confirmed the cognitive status of each participant using the MOCA. However,  
351 we may have missed participants' true cognitive status because of the lack of detailed cognitive  
352 testing and the large time interval since their PET scan. For example, some participants with  
353 preclinical AD may have developed cognitive symptoms since their last PET scan, and some  
354 participants with the clinical label of MCI may have converted to AD. In addition, we did not  
355 establish reliability of ERP in other cognitive domains known to deteriorate in older age, such as  
356 memory and language, and this remains an opportunity for further investigation. Future research  
357 should investigate the added value of cognitive workload measures in the diagnosis, monitoring, and  
358 treatment of individuals at risk of dementia.

## 359 **5 Conclusion**

360 Our current results show that NASA-TLX and ICA are reliable in older adults with and without  
361 cognitive impairment. The lack of strong correlation with P3 ERP measure of cognitive workload  
362 may be due to the multidimensionality of the construct. Further research is needed to understand the  
363 physiological underpinnings of cognitive workload in older adults before these measures can be  
364 considered biomarkers of cognitive decline.

## 365 **6 References**

- 366 1. Cowan N. The Magical Mystery Four: How is Working Memory Capacity Limited, and Why?  
367 *Curr Dir Psychol Sci* 2010;**19**(1):51-57 doi: 10.1177/0963721409359277[published Online  
368 First: Epub Date]].
- 369 2. Kahneman D. *Attention and effort*: Citeseer, 1973.
- 370 3. Bruya B, Tang Y-Y. Is Attention Really Effort? Revisiting Daniel Kahneman's Influential 1973  
371 Book *Attention and Effort*. *Frontiers in Psychology* 2018;**9** doi:  
372 10.3389/fpsyg.2018.01133[published Online First: Epub Date]].

- 373 4. Ranchet M, Morgan JC, Akinwuntan AE, Devos H. Cognitive workload across the spectrum of  
374 cognitive impairments: A systematic review of physiological measures. *Neuroscience &*  
375 *Biobehavioral Reviews* 2017;**80**:516-37 doi: 10.1016/j.neubiorev.2017.07.001[published  
376 Online First: Epub Date]].
- 377 5. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and  
378 Theoretical Research. In: Hancock PA, Meshkati N, eds. *Advances in Psychology*: North-  
379 Holland, 1988:139-83.
- 380 6. Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ. Systematic review of  
381 measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of*  
382 *Surgery* 2018;**105**(5):491-501 doi: 10.1002/bjs.10795[published Online First: Epub Date]].
- 383 7. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors*  
384 *and Ergonomics Society Annual Meeting* 2006;**50**(9):904-08 doi:  
385 10.1177/154193120605000909[published Online First: Epub Date]].
- 386 8. Tubbs-Cooley HL, Mara CA, Carle AC, Gurses AP. The NASA Task Load Index as a measure of  
387 overall workload among neonatal, paediatric and adult intensive care nurses. *Intensive and*  
388 *Critical Care Nursing* 2018;**46**:64-69 doi:  
389 <https://doi.org/10.1016/j.iccn.2018.01.004>[published Online First: Epub Date]].
- 390 9. Devos H, Burns J, Ahmadnezhad P, et al. Reliability of P3 Event-Related Potential during  
391 Working Memory across the Spectrum of Cognitive Aging. *medRxiv*  
392 2020:2020.05.27.20109157 doi: 10.1101/2020.05.27.20109157[published Online First: Epub  
393 Date]].
- 394 10. Ahmadlou M, Adeli A, Bajo R, Adeli H. Complexity of functional connectivity networks in mild  
395 cognitive impairment subjects during a working memory task. *Clinical Neurophysiology*  
396 2014;**125**(4):694-702 doi: <https://doi.org/10.1016/j.clinph.2013.08.033>[published  
397 Online First: Epub Date]].
- 398 11. Galluzzi S, Nicosia F, Geroldi C, et al. Cardiac autonomic dysfunction is associated with white  
399 matter lesions in patients with mild cognitive impairment. *J Gerontol A Biol Sci Med Sci*  
400 2009;**64**(12):1312-5 doi: 10.1093/gerona/glp105[published Online First: Epub Date]].
- 401 12. Granholm EL, Panizzon MS, Elman JA, et al. Pupillary Responses as a Biomarker of Early Risk  
402 for Alzheimer's Disease. *Journal of Alzheimer's Disease* 2017;**56**(4):1419-28 doi:  
403 10.3233/jad-161078[published Online First: Epub Date]].
- 404 13. Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the Pathologic Process in  
405 Alzheimer Disease: Age Categories From 1 to 100 Years. *Journal of Neuropathology &*  
406 *Experimental Neurology* 2011;**70**(11):960-69 doi: 10.1097/nen.0b013e318232a379[published  
407 Online First: Epub Date]].
- 408 14. Wilson RS, Nag S, Boyle PA, et al. Neural reserve, neuronal density in the locus ceruleus, and  
409 cognitive decline. *Neurology* 2013;**80**(13):1202-08 doi:  
410 10.1212/wnl.0b013e3182897103[published Online First: Epub Date]].
- 411 15. Chandler DJ, Jensen P, McCall JG, Pickering AE, Schwarz LA, Totah NK. Redefining  
412 Noradrenergic Neuromodulation of Behavior: Impacts of a Modular Locus Coeruleus  
413 Architecture. *The Journal of Neuroscience* 2019;**39**(42):8239-49 doi: 10.1523/jneurosci.1164-  
414 19.2019[published Online First: Epub Date]].

- 415 16. Samuels E, Szabadi E. Functional Neuroanatomy of the Noradrenergic Locus Coeruleus: Its  
416 Roles in the Regulation of Arousal and Autonomic Function Part I: Principles of Functional  
417 Organisation. *Current Neuropharmacology* 2008;**6**(3):235-53 doi:  
418 10.2174/157015908785777229[published Online First: Epub Date]].
- 419 17. Beatty J, Lucero-Wagoner B. The pupillary system. *Handbook of psychophysiology*, 2nd ed.  
420 New York, NY, US: Cambridge University Press, 2000:142-62.
- 421 18. Kremen WS, Panizzon MS, Elman JA, et al. Pupillary dilation responses as a midlife indicator of  
422 risk for Alzheimer's disease: association with Alzheimer's disease polygenic risk.  
423 *Neurobiology of Aging* 2019;**83**:114-21 doi: 10.1016/j.neurobiolaging.2019.09.001[published  
424 Online First: Epub Date]].
- 425 19. Alnæs D, Sneve MH, Espeseth T, Endestad T, van de Pavert SH, Laeng B. Pupil size signals  
426 mental effort deployed during multiple object tracking and predicts brain activity in the dorsal  
427 attention network and the locus coeruleus. *J Vis* 2014;**14**(4) doi: 10.1167/14.4.1[published  
428 Online First: Epub Date]].
- 429 20. Mathur A, Gehrman J, Atchison DA. Pupil shape as viewed along the horizontal visual field.  
430 *Journal of Vision* 2013;**13**(6):3-3 doi: 10.1167/13.6.3[published Online First: Epub Date]].
- 431 21. Granholm EL, Panizzon MS, Elman JA, et al. Pupillary Responses as a Biomarker of Early Risk  
432 for Alzheimer's Disease. *J Alzheimers Dis* 2017;**56**(4):1419-28 doi: 10.3233/jad-  
433 161078[published Online First: Epub Date]].
- 434 22. Marshall SP. The Index of Cognitive Activity: measuring cognitive workload. *Proceedings of the*  
435 *IEEE 7th Conference on Human Factors and Power Plants* 2002:7-7
- 436 23. *The Index of Pupillary Activity* 2018. ACM Press.
- 437 24. Stark L, Campbell FW, Atwood J. Pupil Unrest: An Example of Noise in a Biological  
438 Servomechanism. *Nature* 1958;**182**(4639):857-58 doi: 10.1038/182857a0[published Online  
439 First: Epub Date]].
- 440 25. Demberg V, Sayeed A. The Frequency of Rapid Pupil Dilations as a Measure of Linguistic  
441 Processing Difficulty. *PLOS ONE* 2016;**11**(1):e0146194 doi:  
442 10.1371/journal.pone.0146194[published Online First: Epub Date]].
- 443 26. Devos H, Akinwuntan AE, Alissa N, Morohunfola B, Lynch S. Cognitive performance and  
444 cognitive workload in multiple sclerosis: Two different constructs of cognitive functioning?  
445 *Mult Scler Relat Disord* 2020;**38**:101505 doi: 10.1016/j.msard.2019.101505[published Online  
446 First: Epub Date]].
- 447 27. Kahya M, Moon S, Lyons KE, Pahwa R, Akinwuntan AE, Devos H. Pupillary Response to  
448 Cognitive Demand in Parkinson's Disease: A Pilot Study. *Front Aging Neurosci* 2018;**10**:90  
449 doi: 10.3389/fnagi.2018.00090[published Online First: Epub Date]].
- 450 28. Moon S, Kahya M, Lyons KE, Pahwa R, Akinwuntan AE, Devos H. Cognitive workload during  
451 verbal abstract reasoning in Parkinson's disease: a pilot study. *Int J Neurosci* 2020:1-7 doi:  
452 10.1080/00207454.2020.1746309[published Online First: Epub Date]].
- 453 29. Myers JS, Alissa N, Mitchell M, et al. Pilot Feasibility Study Examining Pupillary Response  
454 During Driving Simulation as a Measure of Cognitive Load in Breast Cancer Survivors.  
455 *Oncol Nurs Forum* 2020;**47**(2):203-12 doi: 10.1188/20.Onf.203-212[published Online First:  
456 Epub Date]].

- 457 30. Myers JS, Kahya M, Mitchell M, et al. Pupillary response: cognitive effort for breast cancer  
458 survivors. *Support Care Cancer* 2019;**27**(3):1121-28 doi: 10.1007/s00520-018-4401-  
459 0[published Online First: Epub Date]].
- 460 31. Ranchet M, Orlosky J, Morgan J, Qadir S, Akinwuntan AE, Devos H. Pupillary response to  
461 cognitive workload during saccadic tasks in Parkinson's disease. *Behav Brain Res*  
462 2017;**327**:162-66 doi: 10.1016/j.bbr.2017.03.043[published Online First: Epub Date]].
- 463 32. Wang C, Gao J, Li M, et al. Association of cognitive impairment and mood disorder with event-  
464 related potential P300 in patients with cerebral small vessel diseases. *Neuro Endocrinol Lett*  
465 2019;**40**(7-8):333-41
- 466 33. Jervis BW, Bigan C, Jervis MW, Besleaga M. New-Onset Alzheimer's Disease and Normal  
467 Subjects 100% Differentiated by P300. *American Journal of Alzheimer's Disease & Other*  
468 *Dementias* 2019;**34**(5):308-13 doi: 10.1177/1533317519828101[published Online First:  
469 Epub Date]].
- 470 34. Ghani U, Signal N, Niazi IK, Taylor D. ERP based measures of cognitive workload: A review.  
471 *Neuroscience & Biobehavioral Reviews* 2020;**118**:18-26 doi:  
472 10.1016/j.neubiorev.2020.07.020[published Online First: Epub Date]].
- 473 35. Nieuwenhuis S, De Geus EJ, Aston-Jones G. The anatomical and functional relationship between  
474 the P3 and autonomic components of the orienting response. *Psychophysiology*  
475 2011;**48**(2):162-75 doi: 10.1111/j.1469-8986.2010.01057.x[published Online First: Epub  
476 Date]].
- 477 36. Murphy PR, Robertson IH, Balsters JH, O'Connell RG. Pupillometry and P3 index the locus  
478 coeruleus-noradrenergic arousal function in humans. *Psychophysiology* 2011;**48**(11):1532-43  
479 doi: 10.1111/j.1469-8986.2011.01226.x[published Online First: Epub Date]].
- 480 37. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's  
481 disease: recommendations from the National Institute on Aging-Alzheimer's Association  
482 workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*  
483 2011;**7**(3):280-92 doi: 10.1016/j.jalz.2011.03.003[published Online First: Epub Date]].
- 484 38. Vidoni ED, Yeh H-W, Morris JK, et al. Cerebral  $\beta$ -Amyloid Angiopathy Is Associated with  
485 Earlier Dementia Onset in Alzheimer's Disease. *Neurodegenerative Diseases* 2016;**16**(3-  
486 4):218-24 doi: 10.1159/000441919[published Online First: Epub Date]].
- 487 39. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: A  
488 brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005;**53**(4):695-99 doi:  
489 10.1111/j.1532-5415.2005.53221.x[published Online First: Epub Date]].
- 490 40. Bopp KL, Verhaeghen P. Aging and n-Back Performance: A Meta-Analysis. *The Journals of*  
491 *Gerontology: Series B* 2018 doi: 10.1093/geronb/gby024[published Online First: Epub Date]].
- 492 41. Miller KM, Price CC, Okun MS, Montijo H, Bowers D. Is the N-Back Task a Valid  
493 Neuropsychological Measure for Assessing Working Memory? *Archives of Clinical*  
494 *Neuropsychology* 2009;**24**(7):711-17 doi: 10.1093/arclin/acp063[published Online First:  
495 Epub Date]].
- 496 42. Gevins A, Smith ME, McEvoy LK, et al. A cognitive and neurophysiological test of change from  
497 an individual's baseline. *Clinical Neurophysiology* 2011;**122**(1):114-20 doi:  
498 10.1016/j.clinph.2010.06.010[published Online First: Epub Date]].

## Psychometric Properties of Cognitive Workload in Older Adults

- 499 43. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical  
500 and theoretical research. *Advances in psychology* 1988;**52**:139-83
- 501 44. Ikuma LH, Nussbaum MA, Babski-Reeves KL. Reliability of physiological and subjective  
502 responses to physical and psychosocial exposures during a simulated manufacturing task.  
503 *International Journal of Industrial Ergonomics* 2009;**39**(5):813-20 doi:  
504 <https://doi.org/10.1016/j.ergon.2009.02.005>[published Online First: Epub Date]].
- 505 45. Vogels J, Demberg V, Kray J. The Index of Cognitive Activity as a Measure of Cognitive  
506 Processing Load in Dual Task Settings. *Frontiers in Psychology* 2018;**9**(2276) doi:  
507 10.3389/fpsyg.2018.02276[published Online First: Epub Date]].
- 508 46. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG  
509 dynamics including independent component analysis. *J Neurosci Methods* 2004;**134**(1):9-21  
510 doi: 10.1016/j.jneumeth.2003.10.009[published Online First: Epub Date]].
- 511 47. Lopez-Calderon J, Luck SJ. ERPLAB: an open-source toolbox for the analysis of event-related  
512 potentials. *Frontiers in Human Neuroscience* 2014;**8**(213) doi:  
513 10.3389/fnhum.2014.00213[published Online First: Epub Date]].
- 514 48. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*  
515 1979;**86**(2):420-8 doi: 10.1037//0033-2909.86.2.420[published Online First: Epub Date]].
- 516 49. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized  
517 assessment instruments in psychology. *Psychological assessment* 1994;**6**(4):284
- 518 50. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of  
519 clinical measurement. *Lancet (London, England)* 1986;**1**(8476):307-10
- 520 51. Donoghue D, Stokes E. How much change is true change? The minimum detectable change of  
521 the Berg Balance Scale in elderly people. *Journal of Rehabilitation Medicine* 2009;**41**(5):343-  
522 46 doi: 10.2340/16501977-0337[published Online First: Epub Date]].
- 523 52. Xiao YM, Wang ZM, Wang MZ, Lan YJ. [The appraisal of reliability and validity of subjective  
524 workload assessment technique and NASA-task load index]. *Zhonghua Lao Dong Wei Sheng*  
525 *Zhi Ye Bing Za Zhi* 2005;**23**(3):178-81
- 526 53. Lew HL, Gray M, Poole JH. Temporal Stability of Auditory Event-Related Potentials in Healthy  
527 Individuals and Patients With Traumatic Brain Injury. *Journal of Clinical Neurophysiology*  
528 2007;**24**(5):392-97 doi: 10.1097/wnp.0b013e31814a56e3[published Online First: Epub  
529 Date]].
- 530 54. Kahya M, Lyons KE, Pahwa R, Akinwuntan AE, He J, Devos H. Reliability and Validity of  
531 Pupillary Response during Dual-task Balance in Parkinson's Disease. *Archives of Physical*  
532 *Medicine and Rehabilitation* 2020 doi: 10.1016/j.apmr.2020.08.008[published Online First:  
533 Epub Date]].
- 534 55. Kahya M, Liao K, Gustafson K, Akinwuntan A, Devos H. Validation of Pupillary Response  
535 Against EEG during Dual-Tasking Postural Control. *Archives of Physical Medicine and*  
536 *Rehabilitation* 2019;**100**(10):e142 doi: 10.1016/j.apmr.2019.08.434[published Online First:  
537 Epub Date]].
- 538 56. Schmiedek F, Li S-C, Lindenberger U. Interference and facilitation in spatial working memory:  
539 Age-associated differences in lure effects in the n-back paradigm. *Psychology and Aging*  
540 2009;**24**(1):203-10 doi: 10.1037/a0014685[published Online First: Epub Date]].

541 57. Kim M, Barrett AM, Heilman KM. Lateral Asymmetries of Pupillary Responses. *Cortex*  
542 1998;**34**(5):753-62 doi: 10.1016/s0010-9452(08)70778-0[published Online First: Epub Date]].

543

544 **7 Tables**

545 **8 Data Availability**

546 The raw data supporting the conclusions of this article will be made available by the authors, without  
547 undue reservation.

548 **9 Ethics Statement**

549 The studies involving human participants were reviewed and approved by University of Kansas  
550 Medical Center Internal Review Board. The patients/participants provided their written informed  
551 consent to participate Each participant received \$100 for participating in this study.

552 **10 Conflict of Interest**

553 The authors declare that the research was conducted in the absence of any commercial or financial  
554 relationships that could be construed as a potential conflict of interest.

555 **11 Author Contributions**

556 HD, JB, JM, WMB, and KG conceptualized the study. HD, KL, and KG worked out the EEG data  
557 processing steps. HD, PA, and KL administered the tests. HD and JM analysed the data. HD wrote  
558 the initial manuscript. JB, KL, PA, JM, WMB, and KG reviewed the manuscript and provided  
559 valuable comments.

560 **12 Funding**

561 Research reported in this publication was supported by the National Institute on Aging of the  
562 National Institutes of Health under Award Number K01 AG058785. This study was supported in part  
563 by a pilot grant of the KU Alzheimer Disease Center (P30 AG035982). The Hoglund Biomedical  
564 Imaging Center is supported in part by S10 RR29577 and generous gifts from Forrest and Sally  
565 Hoglund. The content is solely the responsibility of the authors and does not necessarily represent the  
566 official views of the National Institutes of Health.

567 **13 Acknowledgments**

568 The authors thank the volunteers for their time and willingness to participate in this research. We are  
569 also grateful for the staff at the KU Alzheimer Disease Center.

570 **14 Figure Legends**

571 Figure 1. Grand Average Event-related Potential Waveform at Fz of (a) 0-back, (b) 1-back and (c) 2-  
572 back.

## Psychometric Properties of Cognitive Workload in Older Adults

573 Figure 2. Bland Altman Plots of (a) 0-back Fz Peak Amplitude (b) 1-back Fz Peak Amplitude; (c) 2-  
574 back Fz Peak Amplitude; (d) 0-back Fz Peak Latency; (e) 1-back Peak Latency; (f) 2-back Peak  
575 Latency.