Evaluation of the validity of the Psychology Experiment Building Language tests of vigilance, auditory memory, and decision making

Brian Piper, Shane T Mueller, Sara Talebzadeh, Min Jung Ki

Background. The Psychology Experimental Building Language (PEBL) http://pebl.sourceforge.net/ test battery is a popular application for neurobehavioral investigations. This study evaluated the correspondence between the PEBL and the non-PEBL versions of four executive function tests.

Methods. In one cohort, young-adults (N = 44) completed both the Conner's Continuous Performance Test ($_c$CPT) and the PEBL CPT ($_p$CPT) with the order counter-balanced. In a second cohort, participants (N = 47) completed a non-computerized (Wechsler) and a computerized (PEBL) Digit Span ($_w$DS or $_p$DS) both Forward and Backward. Participants also completed the Psychological Assessment Resources or the PEBL versions of the Iowa Gambling Task ($_{PAR}$IGT or $_{PEBL}$IGT).

 Results. The between test correlations were moderately high (reaction time r = 0.78, omission errors r = 0.65, commission errors r = 0.66) on the CPT. DS Forward was significantly greater than DS Backward independent of the test modality. The total $_w$DS score was moderately correlated with the $_p$DS (r = 0.56). The $_{PAR}$IGT and the $_{PEBL}$IGTs showed a very similar pattern for response times across blocks, development of preference for Advantageous over Disadvantageous Decks, and Deck selections. However, the amount of money earned (score – loan) was significantly higher in the $_{PEBL}$IGT during the last Block.

Conclusions. These findings are broadly supportive of the criterion validity of the PEBL measures of sustained attention, short-term memory, and decision making. Select differences between workalike versions of the same test highlight how detailed aspects of implementation may have more important consequences for computerized testing than has been previously acknowledged.

1   **Evaluation of the Validity of the Psychology Experiment Building Language Tests of**

2   **Vigilance, Auditory Memory, and Decision Making**

3
4   Brian J. Piper[abc], Shane T. Mueller[d,] Sara Talebzadeh[b], & Min Jung Ki[b]

5

6   [a]Bowdoin College, Brunswick, ME 04011 USA

7   [b]Husson University, Bangor, ME 04401 USA

8   [c]Willamette University, Salem, OR 97301 USA

9   [d]Michigan Technological University, Houghton, MI 49931 USA

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33   Corresponding Author:
34   Brian J. Piper, Ph.D.
35   255 Main Street
36   Department of Psychology & Neuroscience Program
37   Bowdoin College
38   Brunswick, ME 04001
39   psy391@gmail.com; bpiper@bowdoin.edu
40

41   **Abstract**
42
43

44   **Background.** The Psychology Experimental Building Language (PEBL)

45   http://pebl.sourceforge.net/ test battery is a popular application for neurobehavioral

46   investigations. This study evaluated the correspondence between the PEBL and the non-PEBL

47   versions of four executive function tests.

48   **Methods.** In one cohort, young-adults (N = 44) completed both the Conner's Continuous

49   Performance Test ($_c$CPT) and the PEBL CPT ($_p$CPT) with the order counter-balanced. In a

50   second cohort, participants (N = 47) completed a non-computerized (Wechsler) and a

51   computerized (PEBL) Digit Span ($_w$DS or $_p$DS) both Forward and Backward. Participants also

52   completed the Psychological Assessment Resources or the PEBL versions of the Iowa Gambling

53   Task ($_{PAR}$IGT or $_{PEBL}$IGT).

54   **Results**. The between test correlations were moderately high (reaction time $r = 0.78$, omission

55   errors $r = 0.65$, commission errors $r = 0.66$) on the CPT. DS Forward was significantly greater

56   than DS Backward independent of the test modality. The total $_w$DS score was moderately

57   correlated with the $_p$DS ($r = 0.56$). The $_{PAR}$IGT and the $_{PEBL}$IGTs showed a very similar pattern

58   for response times across blocks, development of preference for Advantageous over

59   Disadvantageous Decks, and Deck selections. However, the amount of money earned (score −

60   loan) was significantly higher in the $_{PEBL}$IGT during the last Block.

61   **Conclusions**. These findings are broadly supportive of the criterion validity of the PEBL

62   measures of sustained attention, short-term memory, and decision making. Select differences

63   between workalike versions of the same test highlight how detailed aspects of implementation

64   may have more important consequences for computerized testing than has been previously

65   acknowledged.

INTRODUCTION

An increasingly large collection (>100) of classic clinical psychology and behavioral neurology

tests have been computerized and made freely available (http://pebl.sf.net) over the past decade.

This Psychology Experiment Building Language (PEBL) test battery has been downloaded over

20,000 times/year (Mueller, 2012, 2015; Mueller & Piper, 2014) and use continues to increase

(Fox et al. 2013; Lipnicki et al., 2009a, 2009b; Piper, 2010). The PEBL tests have been

employed in studies of traumatic brain injury (Danckert et al., 2011), behavioral pharmacology

(Aggarwal et al., 2011; Lyvers & Tobias-Webb, 2010), aging (Clark & Kar, 2011; Piper et al.

2012), Parkinson's disease (Peterson, et al., 2015) and behavioral genetics (Wardle et al. 2013;

González-Giraldo et al., 2014) by investigators in developed and developing countries and the

tests have been administered in many languages. A key step in PEBL battery development is to

evaluate criterion validity (i.e., the extent to which its dependent measures predict other existing

measures) by determining whether performance on PEBL tests is similar to the established

versions of the tests. Although the PEBL tests were developed based on the method sections of

the peer reviewed literature, this direct comparison is important because some potentially

important procedural details may have been omitted, described ambiguously, or misinterpreted.

Four tests were selected for the present report: the PEBL Continuous Performance Test

(CPT), Digit Span Forward (DS-F), DS Backward (DS-B), and the Iowa Gambling Task (IGT).

These tests were chosen because they assess theoretically important constructs (vigilance,

attentional capacity, short-term memory, and decision making), have an extensive history, and

their neural substrates have been examined in lesion and neuroimaging studies. Each of these

measures is described in more detail below.

Continuous Performance Test (CPT)

90   CPTs have an extensive history and exist in multiple forms (Mackworth, 1948; Rosvold,

91   Mirsky et al., 1956; Anderson et al., 1969; Earle-Boyer et al., 1991; Greenberg & Waldman,

92   1993; Dougherty, Marsh, & Mathias, 2002; Riccio et al., 2002). These tests require participants

93   to maintain vigilance and respond to the presence of a specific stimulus within a set of

94   continuously presented distracters. A key quality of a CPT is that, rather than being a series of

95   trials that each require a response; a CPT is presented as a continuous series of stimuli whose

96   timing does not appear to depend on the speed or presence of a response, and so it represents a

97   continuous mental workload that has been used to assess vigilance, alertness, attention, and

98   related concepts. The CPT, version II, of Keith Conners, PhD (hence-forth $_c$CPT) has been

99   widely used as a neuropsychological instrument to measure attention in children and adults

100   (Conners, 2004; Piper, et al., 2010, 2011). The fourteen minute $_c$CPT involves responding to

101   target letters (letters A – S presented for 1, 2, or 4 sec each) and inhibiting responses to foils (the

102   letter X). Dependent measures include response times (RT), the variability of RT, the absence of

103   response to target stimuli (omission errors), and responses to the foil (commission errors). There

104   is some debate regarding the utility of the $_c$CPT to aid in a diagnosis of Attention Deficit

105   Hyperactivity Disorder (ADHD) (Cohen & Shapiro, 2007; McGee, Clark, & Symons, 2000).

106   Overall, the strengths of this instrument are its objectivity, simplicity, brevity, a sizable

107   normative sample (Conners & Jeff, 1999; Homack & Riccio, 2006), and it has been shown to be

108   sensitive to psychostimulants used to treat attention disorders (Solanto et al., 2009). In addition,

109   the neural substrates of vigilance have been characterized and involve a network that includes the

110   prefrontal, frontal, and parietal cortex and the striatum (Ogg et al., 2008; Riccio, et al., 2002).

111   *Digit Span Forward and Backward (DS-F and DS-B)*

112   DS type tests are found in the Wechsler assessments as well as in other

113 neuropsychological batteries. A string of numbers are presented (e.g. 7, 1, 6 at a rate of one digit

114 per second) and the participant either repeats them in the same (DS-F) or the reverse (DS-B)

115 sequence. Although DS-F and DS-B are procedurally similar, and they are sometimes viewed as

116 simple short-term memory tasks (St. Clair-Thompson & Allen, 2013), the former is sometimes

117 treated as a measure of 'pure storage' whereas the latter is viewed as involving more executive

118 control and thus considered a "working memory" task (Lezak et al., 2012). DS-B induces

119 greater activity in the prefrontal cortex than DS-F (Keneko et al., 2011).

120   A direct comparison of DS by mode of administration revealed lower DS Forward and

121 Backward when completed over the telephone with voice recognition as compared to in-person

122 administration (Miller et al., 2013). However, a moderate correlation (r = .53) in DS total was

123 identified with traditional and computerized administration (Paul et al., 2005).

124 *Iowa Gambling Test (IGT)*
125
126   Antoine Bechara, PhD and colleagues at the University of Iowa College of Medicine

127 developed a novel task to quantify abnormalities in decision making abilities. Originally, what

128 became known as the Iowa Gambling Task (IGT) involved selecting cards from four physical

129 decks of cards. Each deck had a different probability of wins versus losses. Two decks are

130 Disadvantageous and two are Advantageous, because some deck selections will lead to losses

131 over the long run, and others will lead to gains. Neurologically intact participants were reported

132 to make the majority (70%) of one-hundred selections from the Advantageous (C & D) decks. In

133 contrast, patients with lesions of the prefrontal cortex showed the reverse pattern with a strong

134 preference for the Disadvantageous (A & B) decks (Bechara et al., 1994, although see Buelow &

135 Suhr, 2009; Steingroever, et al., 2013). However, another research team, employing a gambling

136 task that they programmed, determined that college-aged adults showed a response pattern that is

137 very similar to patients with frontal lesions (Caroselli, et al., 2006). IGT type tasks have become

138 increasingly popular for research purposes to examine individual differences in decision making

139 including in pathological gamblers, substance abusers, Attention Deficit Hyperactivity Disorder

140 (ADHD), and in other neurobehavioral disorders (Buelow & Suhr, 2009; Verdejo-Garcia, et al.,

141 2007). One key characteristic of the IGT is that there is substantial carryover of learning with

142 repeated administrations in normal participants (Bechara, et al., 2000a; Fernie & Tumney, 2006;

143 Verdejo-Garcia et al. 2007). Bechara, in conjunction with Psychological Assessment Resources

144 (PAR), distributes a computerized version of the IGT (Bechara, 2007). The IGT is also one of

145 the more widely employed tests in the PEBL battery (Hawthorne et al., 2011; Lipnicki, et al.,

146 2009a, 2009b; Mueller & Piper, 2014) and so itself has been used in many different contexts.

147 Many variations on IGT procedures have been developed over the past two-decades. The

148 $_{PEBL}$IGT employs consistent rewards and punishment (e.g. -$1,250 for each selection from Deck

149 B) as described by Bechara et al. 1994. The $_{PAR}$IGT utilizes the ascending schedule of rewards

150 and punishments (e.g. -$1,250 for early deck selections and decreasing by $250 increments)

151 (Bechara et al. 2000b).

152   The primary objective of this report was to determine the similarity between the PEBL

153 and non-PEBL versions of executive function measures. Where applicable, intra-test correlations

154 were also examined as this is one criteria used to evaluate test equivalence (Bartram, 1994).

155 Although not specified a priori, the IGT dataset was also used to critically examine the

156 sensitivity of the IGT to identify clinically meaningful individual differences in decision making

157 abilities.

159

MATERIALS & METHODS

161   Participants. The participants (N = 44; Age = 18-24, Mean = 18.7 ± 0.2; 68.2% female;

162   23.9% non-white; 7.3% ADHD) were college students receiving course credit in the CPT study.

163   A separate cohort (N = 47; Age = 18-34, Mean = 18.8 ± 0.3; 59.6% female; 14.9% non-white;

164   10.6% ADHD) of college students completed the DS/IGT study.

165   Procedures. All procedures were approved by the Institutional Review Board of

166   Willamette University (first cohort) or the University of Maine, Orono (second cohort).

167   Participants were tested individually with an experimenter in the same room. Each participant

168   completed an informed consent and a short demographic form which included items about sex,

169   age, whether they had been diagnosed by a medical professional with ADHD. Next, the first

170   cohort completed either the $_p$CPT (modified from the default in PEBL version 0.11) or Version II

171   of the $_c$CPT, including the two-minute practice trial, with the order counter-balanced on desktop

172   computers running Windows and not connected to the internet. As data collection for each CPT

173   takes 14 minutes and is intentionally monotonous, the PEBL Tower of London (Piper et al.

174   2012) was completed between each CPT as a brief ($\approx$5 min) distractor task. The instructions of

175   the $_p$CPT were:

176   You are about to take part in an experiment that involves paying attention to letters on a
177   screen. It will take about 14 minutes. You will see letters presented on a screen quickly.
178   Your goal is to press the spacebar as fast as possible after each letter, except if the letter
179   is an 'X'. DO NOT RESPOND to X stimuli.
180
181   A total of 324 target letters (A, B, C, D, E, F, G, H, I, J, K, L, M, O, P, Q, R, S, U) and 36

182   foils (X) were presented with an inter-stimulus interval of 1, 2, or 4 seconds. The primary

183   dependent measures were the RT on correct trials in ms, the standard deviation (SD) of RT,

184 omission and commission errors. The $_p$CPT source code is also included in the supplementary

185 materials.

186   The second cohort completed a short demographic form (described above) followed by

187 the PEBL and non-PEBL tasks (DS-F, DS-B, and IGT) with the order counterbalanced across

188 testing sessions. PEBL, version 0.14, was installed on Dell laptops. Both laptops were connected

189 to Dell touchscreen monitors which were used for selecting responses on the IGT.

190   The Wechsler DS ($_w$DS) consists of two trials for each number of items each read aloud

191 by the experimenter at a rate of one per second beginning with two items. Discontinuation

192 occurred when both trials for a single number of items were answered incorrectly. The maximum

193 total score for DS Forward and Backward is sixteen and fourteen, respectively. The PEBL Digit

194 Span ($_p$DS) source code was modified slightly from the default version so that stimuli were

195 presented via headphones (one per 1,000 ms) but not visually. Two trials were completed for

196 each number of items starting with three items. Digit stimuli were generated randomly such that

197 each sequence contained no more than one of each digit. Discontinuation occurred when both

198 trials for a single number of items were answered incorrectly. An important methodological

199 difference between the $_w$DS and the $_p$DS involves how responses are collected. The traditional

200 $_w$DS involves oral responses coded by the experimenter. The $_p$DS involves typed input with the

201 response sequence visible on-screen as it is made. Furthermore, blank entries are permitted and

202 participants have the ability to delete erroneous responses (see supplemental materials for the

203 source code and task instructions).

204   The $_{PAR}$IGT (Version 1.00) was installed on a laptop (Dell Latitude E6410) with

205 headphones. The administration instructions were shown and read/paraphrased for the participant

206 (Bechara et al. 2000a, Bechara, 2007) and the default settings were used. The $_{PEBL}$IGT was also

207 administered with the order counterbalanced. Due to pronounced practice effects with the IGT

208 (Bechara et al. 2000a; Birkett et al., 2015; Verdejo-Garcia et al. 2007), only data from the IGT

209 administered first was examined. The $_{PEBL}$IGT has modifications contributed by P. N. Bull

210 (supplemental materials) and is a more refined version of the task than has been used previously

211 (Hawthorne et al., 2011; Lipnicki et al., 2009a, 2009b). If scores go below zero, participants will

212 receive a second $2,000 loan. Importantly, the $_{PEBL}$IGT is based on the procedures described in

213 Bechara et al. 1994 while the $_{PAR}$IGT is based on those described in great detail in Bechara et al.

214 2000b. The instructions are 14% shorter on the $_{PEBL}$IGT but perhaps the largest procedural

215 difference is the negative consequences of Disadvantageous Decks are amplified in the $_{PAR}$IGT

216 (Table 1).

217   <u>Statistical analyses</u>: The $_{P}$CPT output text files were imported into Excel and all analyses

218 were subsequently conducted using Systat, version 13.0. The distribution on some measures

219 (e.g. RT), were, as anticipated, non-normal, therefore both Pearson (r_P) and Spearman rho (r_S)

220 correlation coefficients were completed. As the $_{P}$CPT default settings express the variability in

221 RT slightly differently (SD) than the cCPT (SE), the PEBL output was converted to the SE

222 according to the formula $SD/(N – 1)^{0.5}$ where N is the total number of correct trials across the

223 three inter-trial intervals. Differences in correlations between the $_{P}$CPT and $_{C}$CPT were evaluated

224 with a Fisher r to Z transformation (http://vassarstats.net/rdiff.html). As the $_{W}$DS starts at an

225 easier level (2 digits) than the $_{P}$DS (3 digits), two additional points were added to each (Forward

226 and Backward) $_{P}$DS for comparison purposes. The primary dependent measure on the IGT was

227 Deck selections but Response Times on each Block of twenty-trials and the compensation (score

228 minus loan) for each trial was also documented. The NET was calculated as Advantageous

229   minus Disadvantageous Deck selections. Mean data are presented with the SEM and $p < .05$

230   considered statistically significant.

232

233 **RESULTS**

234 *CPT*

235   Substantial individual differences in sustained attention were observed in this sample.

236 The percentiles for each $_c$CPT measure are shown in Table 2.

237   The inter-test correlations were generally satisfactory. The correlation was excellent for

238 reaction time ($r_P(42) = +.78$, $r_S(42) = +.80$, $p < .0005$, Figure 1A). Reaction time variability was

239 also moderately high ($r_P(42) = +.66$, $r_S(42) = +.27$, $p < .0005$) but this association should be

240 viewed with caution as removal of one extreme score (15.9, 23.3) reduced this correlation

241 considerably ($r_P(42) = +.20$, $p = .19$; data not shown). Omission errors ($r_P(42) = +.65$, $p < .0005$,

242 $r_S(42) = +.31$, $p < .05$) and commission errors ($r_P(42) = +.66$, $r_S(42) = +.66$, $p < .0005$) showed

243 good correlations across tests (Figure 1B & 1C).

244   Mean reaction time on correct trials differed slightly (by 12 ms) between tests, which was

245 statistically significant ($_c$CPT = 327.1 ± 6.5, $_p$CPT = 315.2 ± 4.7, $t(43) = 2.91$, $p < .01$). The

246 difference in the SE of RT was clearly different ($_c$CPT = 5.3 ± 0.4, $_p$CPT = 3.3 ± 0.5, $t(43) =$

247 5.60, $p < .0005$) but there was no difference for omission errors ($_c$CPT = 2.6 ± 0.6, $_p$CPT = $2.3 \pm$

248 0.7, $t(43) = 0.51$, $p = .61$) or commission errors ($_c$CPT = 18.1 ± 1.1, $_p$CPT = 17.3 ± 1.0, $t(43) =$

249 0.96, $p = .34$).

250   An analysis of the intra-test Spearman correlations among the variables of each test was

251 also conducted (Table 3). Several significant correlations were identified. However, with the

252 exception of a trend for the RT SE ($p = .055$), the correlations did not differ across tests.

253 *DS*

254   Figure 2A shows the anticipated higher score for Forward (10.0 ± 0.3, Min = 6, Max =

255   13) relative to Backward (6.3 ± 0.3, Min = 3, Max = 11) on the $_w$DS. The correlation between

256   Forward and Backward was moderate ($r_P(45) = .43$, $p < .005$; $r_S(45) = .41$, $p < .005$).

257   Figure 2A also depicts an elevated score for Forward (10.5 ± 0.4, Min = 3, Max = 15)

258   compared to Backward (8.2 ± 0.3, Min = 4, Max = 12, $t(46) = 5.10$, $p < .0005$) for the $_p$DS. The

259   correlation between Forward and Backward was not significant ($r_P(45) = .22$, $p > .10$; $r_S(45) =$

260    $.28$, $p = .054$). The $_p$DS-B was significantly higher than $_w$DS-B ($t(46) = 6.43$, $p < .0005$).

261   The correlation between computerized and non-computerized DS was intermediate for

262   Forward ($r_P(45) = .42$, $p < .005$; $r_S(45) = .45$, $p < .005$) and Backward ($r_P(45) = .49$, $p < .001$;

263    $r_S(45) = .467$, $p < .001$). Figure 2B shows the association between the DS total (Forward +

264   Backward) across test modalities was moderate ($r_S(47) = .51$, $p < .0005$).

265

266   *IGT*

267   The NET 1 to 5 percentile score was 38.0 ± 4.4 (Min = 5, Max = 90) on the $_{PAR}$IGT. The

268   standardized (T_{50}) score was 47.2 ± 1.5 (Min = 34.0, Max = 63.0) which was non-significantly

269   lower than the normative mean of 50 (one sample $t(23) = 1.91$, $p = .069$). Response Times

270   showed a clear decrease over the course of the session with shorter times on Block 2 ($t(23) =$

271   4.49, $p < .0005$), Block 3 ($t(23) = 5.93$, $p < .0005$), Block 4 ($t(23) = 5.42$, $p < .0005$) and Block 5

272   ($t(23) = 5.07$, $p < .0005$) relative to Block 1 (Figure 3A). Responses on the first Block showed a

273   trend favoring Disadvantageous over Advantageous Decks ($t(23) = 1.90$, $p = .07$) with the

274   reverse pattern on the last Block (Figure 3C). Similarly, there was a trend toward greater

275   Advantageous selections on Block 5 (11.0 ± 0.9) compared to Block 1 ($t(23) = 1.83$, $p = .081$).

276   Across all Blocks, participants made fewer selections from Deck A' compared to Deck B' ($t(23)$

277   = 8.98, p < .0005), Deck C' (t(23) = 3.48, $p \leq$ .002) or Deck D' (t(23) = 3.65, $p \leq$ .001).

278   Participants made more selections from Deck B' compared to Deck C' (t(23) = 2.79, $p \leq$ .01) or

279   Deck D' (t(23) = 2.72, p < .02, Figure 3E). Almost half (45.8%) of participants made more

280   selections from Disadvantageous (C' + D') than Advantageous (C' + D') Decks. Figure 4A

281   shows the Deck selections on each trial for a participant with the median NET 1 to 5. Half

282   (50.0%) of participants received the second $2,000 loan. The amount earned (score minus loan)

283   increased during the Block 1, dropped below zero during Block 3, and was negative by test

284   completion (-$1,099.58 $\pm$ 191.20, Min = -3,015, Max = 1,475, Figure 3G).

285   Relative to the first Block, RTs were significantly shorter on Block 2 (t(18) = 2.85, p <

286   .02), Block 3 (t(18) = 7.45, p < .0005), Block 4 (t(18) = 4.26, $p \leq$ .0005), and Block 5 (t(16) =

287   4.59, p < .0005, Figure 3B) on the $_{PEBL}$IGT. There were more selections from the

288   Disadvantageous than the Advantageous Decks on Block 1 (t(18) = 2.98, p < .01, Figure 3D).

289   When collapsing across the five Blocks, over-two thirds (68.4%) of respondents made more

290   selections from Disadvantageous than Advantageous Decks. Fewer selections were made from

291   Deck A compared to Deck B (t(18) = 4.27, p < .0005) or Deck D (t(18) = 2.45, p < .03).

292   Participants made non-significantly more selections on Deck B compared to Deck C (t(18) =

293   2.05, p = .055, Figure 3F). Figure 4B depicts the Deck selections over the course of the test for a

294   participant with the median NET 1 to 5. Very few (10.5%) participants received the second

295    $2,000 loan. Compensation, defined as the score minus the loan, grew during the Block 1,

296   dropped towards zero in Block 2, and stayed negative for the remainder of the test. The $_{PEBL}$IGT

297   money was significantly lower than $_{PAR}$IGT during trials 16 to 18 and 23 but higher from trial 74

298   until test completion (-$269.74 $\pm$ 255.93, Min = 2,425, Max = 1,950, Figure 3G).

300

DISCUSSION

302   The PEBL software is becoming a ubiquitous tool in the social and biomedical sciences

303 (Mueller & Piper, 2014). Although this widespread use in numerous contexts has helped to

304 establish the general reliability and validity of specific tests, the publication of additional

305 systematic validation studies comparing their results to existing tests will help establish their

306 suitability for use in basic research and clinical neuroscience applications. This report identifies

307 some procedural similarities, and also differences, between the PEBL and commercial versions

308 of ostensibly equivalent tests.

309   ***CPT Tests.*** The CPT developed by Conners and colleagues has been, and will likely

310 continue to be, an important instrument widely employed for applied and research purposes. The

311 mean RT, variability of RT, omission and commission errors are similar to those reported

312 previously with college students as participants (Burton et al., 2010). Moderate to strong

313 correlations across tests were observed on the CPT measures across platforms. The origin of any

314 inter-test differences is multifaceted and could include procedural details (*e.g.* software

315 algorithms), interactions between software and hardware, particularly for RTs (Plant & Quinlan,

316 2013), or participant variance due to repeated testing. Importantly, the inter-test reliability of the

317 $_p$CPT and the $_c$CPT are bound by the test-retest reliability of both measures. Previous research

318 has established moderate to high test-retest reliability for the cCPT, in the same range as our

319 inter-test reliability measures. For example, Conners (2004) reported test-retest correlations of

320 0.55 to 0.84 when the cCPT was administered twice with an inter-test interval of two weeks.

321 Similarly, in a study of twelve children taking the cCPT, Soreni, Crosbie, Ickowicz, and

322 Schachar (2009) found the inter-class correlation coefficients for ommission errors: .09;

323   commission errors:.72; RT: .76; and RTSE of .63. Similarly, Kuntsi et al. (2005), showed for a

324   group of 47 children using a similar go/no-go CPT, inter-class r scores ranged from .7-.88 on RT

325   scores; 0.26-.83 on SD of RT, and .54-.7 on commission errors. Although the experience of the

326   participants was similar when completing the $_C$CPT and the $_P$CPT, some of the algorithms

327   employed in the $_C$CPT are unpublished or could not be verified by the authors. This is

328   particularly a concern for the signal detection measures (Stanislaw & Todorov, 1999) and

329   therefore *d'* and *Beta* were not compared across platforms. Notably, similarity of intra-test

330   correlations is one criterion for the equivalence of measures (Bartram, 1994). The pattern of

331   results with this sample identified in Table 2 generally supports this criterion for the $_P$CPT.

332   ***DS-F and DS-B Tests.*** DS type tasks have an extensive history and have been

333   implemented in an analogous format to the $_W$DS for over a century (Richardson, 2007).

334   Importantly, the test-retest reliability of $_W$DS is moderate (*r* = .68) (Dikmen, Heaton, Grant, &

335   Temkin, 1999). DS-F did not differ between $_W$DS and $_P$DS. Although DS-B was less than DS-F

336   for the $_W$DS and the $_P$DS, the magnitude of reduction was attenuated on the $_P$DS. A subset of

337   participants (≈15%) either were rehearsing the digits aloud or on the keyboard while they were

338   being presented on the $_P$DS. Use of these strategies could change the fundamental nature of the

339   constructs being measured. It is important to emphasize that although stimuli are present aurally

340   for both the $_W$DS and the $_P$DS, response execution is oral for the $_W$DS but typed for the $_P$DS. The

341   format of how stimuli is presented and executed is known to produce detectable differences

342   (Karakas et al., 2002). The correlation between the $_P$DS and the $_W$DS was only moderate. This

343   could be due to modality effects or the use of a college-aged sample may have resulted in a

344   restriction of range which attenuated the associations. In principle, voice recognition algorithms

345   would make $_W$DS and $_P$DS more similar. Other investigators that are refining this technology

346   have identified moderate correlations across modalities (Forward = .48, Backward = .50) but

347   difficulties recognizing the responses of participants with accents is not trivial (Miller et al.

348   2014). More generally, perhaps the notion of the $_{w}$DS as the "gold standard" is questionable.

349   Computerized administration offers the potential of delivering stimuli at a more consistent rate,

350   intensity, and clarity than traditional methods (Woods et al., 2011). The use of more trials per

351   number of digits and alternative procedures for advancement to the difficulty threshold may

352   improve the precision of DS measurement.

353   ***IGT Tests.*** The IGT is sometimes described as a "one-shot" measure of executive

354   function. As such, this investigation did not attempt to evaluate correlations between the $_{PEBL}$IGT

355   and the $_{PAR}$IGT and instead examined response patterns within each test. The $_{PEBL}$IGT and the

356   $_{PAR}$IGT have many procedural similarities but also some differences (Table 1) which may not be

357   widely appreciated. Although there were pronounced individual differences, the $_{PAR}$CPT

358   percentiles were well different than fifty for this collegiate sample. On the primary dependent

359   measure (deck selections), there was a high degree of similarity between the $_{PAR}$IGT and

360   $_{PEBL}$IGT. For example, the development across trials for a preference of Advantageous over

361   Disadvantageous Decks was evident with both tests (Figure 2C & 2D). The choice of individual

362   decks (e.g. Deck B was twice as commonly selected as Deck A) was identified with the $_{PAR}$IGT

363   and the $_{PEBL}$IGT (Figure 2E & 2F). Response times across Blocks were virtually identical in both

364   computerized platforms (Figure 2A & 2B). However, the compensation awarded at the end of the

365   test, a secondary measure (Bechara, 2007), was significantly greater on the $_{PEBL}$IGT. The losses

366   associated with Disadvantageous Decks in the $_{PEBL}$IGT (Deck B = -$1,250) are much less

367   pronounced than those in the $_{PAR}$IGT punishments (Deck B starts at -$1,250 but increases up to -

368   $2,500). Although this procedural difference did not produce other pronounced effects in this

369   sample, future versions of PEBL will allow the experimenter to select among the original (A B C

370   D) IGT (Bechara et al., 1994) or the variant (A' B' C' D') task (Bechara et al., 2000a). Due to

371   this key methodological difference, results from the $_{PEBL}$IGT (Hawthorne, Weatherford, &

372   Tochkov, 2011; Lipnicki, et al., 2009a, 2009b) are unlikely to be identical to what would be

373   obtained if the $_{PAR}$IGT was employed.

374   Although not the principle goal of this study, these datasets provided an opportunity to

375   identify substantial individual differences with both the $_{PAR}$IGT and the $_{PEBL}$IGT. One concern

376   with quantifying decision making with the IGT is that there is considerable heterogeneity of

377   responding, even by normal (i.e. neurologically intact) participants (Steingroever et al., 2013).

378   For example, Carolselli and colleagues determined that over two-thirds (69.5% versus 68.4% in

379   the present study) of university students completing an IGT based on Bechara et al., (1994) made

380   more selections from Disadvantageous than Advantageous Decks (Caroselli et al., 2006). A

381   similar pattern with the $_{PAR}$IGT was also identified in a separate sample with 70.3% of college

382   students from the southwestern U.S. again choosing Disadvantageous over Advantageous Decks

383   (Piper et al., 2015). If forced to choose whether the median participants in this college student

384   sample (Figure 4) show a response pattern more similar to the typical control or to a patient

385   (EVR 318) from Bechara et al. 1994, we would select the lesioned profile. Similarly, Bechara

386   and colleagues noted that over one-third (37%) of controls fell within the range of ventromedial

387   prefrontal lesion group when using the ascending (A' B' C' D') paradigm (Bechara & Damasio,

388   2002). Findings like this, as well as the present outcomes (i.e. almost half favoring the

389   Disadvantageous Decks with the $_{PAR}$IGT) call into question the clinical utility of this test (see

390   also the meta-analysis by Steingroever et al. 2013).

391   The benefit of open-source neurobehavioral tests like the $_{PEBL}$IGT is that the source code

392   is readily available (see supplemental materials) and anyone, independent of their financial

393   resources, can use PEBL. This contributes to the democratization of science. It must also be

394   emphasized that there is substantial room for improved construct validity and test-retest

395   reliability for the IGT (Buelow & Suhr, 2009). Anyone, even with limited computer

396   programming expertise, who is interested in modifying task parameters and generating future

397   generations of decision making paradigms may do so, which, hopefully, will result in tests that

398   have even better psychometric properties. The transparency and flexibility of PEBL are

399   advantages over proprietary computerized neurobehavioral applications. Full disclosure of all

400   methodological information including the underlying programming of computerized

401   neurobehavioral tests is consistent with the dissemination policy of the National Science

402   Foundation (NSF, 2015) and others. However, the modifiability of PEBL is a bit of a double-

403   edged sword in that tasks like the IGT have undergone substantial refinement over the past

404   decade. At a minimum, investigators that make use of PEBL, PAR, or other applications must

405   include information in their methods sections about what version of the software they utilized.

406   One potential limitation of this report is the samples consisted primarily of young adult

407   college students, whereas in clinical settings, these tests are used across the lifespan (children to

408   adult) with a broad range of educational and mental, and psychological profiles. However, a

409   restriction of range for the dependent measures (see Table 2 and the range of the Minimum and

410   Maximum on both $_{PAR}$IGT and $_{W}$DS) does not appear to be an appreciable concern for this

411   dataset, possibly because both cohorts included some individuals with ADHD, including ones

412   not currently taking their stimulant medications. As noted earlier, the characteristics of this

413   convenience sample is more comparable to those employed by others (Caroselli et al. 2006).

414 The PEBL software currently consists of over one-hundred tests of motor function, attention,

415 learning, memory, and executive function in many different languages, and so additional

416 validation studies with more diverse (age, ethnicity, socioeconomic status, computer experience)

417 samples are warranted. Possibly, a second limitation is the few procedural differences between

418 the $_{PAR}$IGT and $_{PEBL}$IGT (Table 1) were not identified until after the data had been collected.

419 Identification of all the essential procedural variables for proprietary measures is not trivial, nor

420 is that even a goal for PEBL test development. Future releases of PEBL (0.15) will however

421 contain an IGT based on the Bechara et al. 2000b as well as other procedural variations.

422 **Conclusions**

423   This report identified a high degree of consistency between the $_{C}$CPT and $_{P}$CPT, the $_{W}$DS

424 and the $_{P}$DS Forward, and the $_{PAR}$IGT and $_{PEBL}$IGT. Further procedural refinements in this open-

425 source software battery will continue to enhance the utility of the PEBL to investigate individual

426 differences in neurocognition.

428

ACKNOWLEDGEMENTS

434

435

436

438

REFERENCES

440   Aggarwal R, Mishra A, Crochet P, Sirimanna P, & Darzi A. 2011. Effect of

441   caffeine and taurine on simulated laparoscopy performed following sleep deprivation.

442   *British Journal of Surgery,* 98:1666-1672. doi: 10.1002/bjs.7600.

443   Anderson VE, Siegel FS, Fisch RO, Wirt RD. 1969. Responses of a Phyenylketonuric children

444   on a continuous performance test. *Journal of Abnormal Psychology,* 74: 358-362.

445   Bartram D. 1994. Computer-based assessment. In C. L. Cooper & I. T. Robertson

446   (Eds.), *International Review of Industrial and Organizational Psychology* (Vol. 9, pp.

447   31–69). Chichester, England: Wiley.

448   Bechara A. 2007. Iowa Gambling Task Professional Manual. Psychology Assessment

449   Resources, Lutz: FL.

450   Bechara A, Damasio AR, Damasio H, Anderson SW. 1994. Insensitivity to future consequences

451   following damage to human prefrontal cortex. *Cognition,* 50: 7-15.

452   Bechara A, Damasio H. 2002. Decision-making and addiction (part I): Impaired activation of

453   somatic states in substance dependent individuals when pondering decisions with

454   negative future consequences. *Neuropsychologia,* 40: 1675-1689.

455   Bechara A, Damasio H, Damasio AR. 2000a. Emotion, decision making and orbitofrontal

456   cortex. *Cerebral Cortex,* 10: 295-307.

457   Bechara A, Tranel D, Damasio, H. (2000b). Characterization of the decision-making

458   deficit of patients with ventromedial prefrontal cortex lesions. *Brain,* 123: 2189-2202.

459   Birkett MB, Averett A, Soni J, Piper B. 2015. The influence of visual and auditory cue salience

460   on Iowa Gambling Task performance. Poster presented at the American Psychological

461   Association.

462 Buelow MT, Surh JA. 2009. Construct validity of the Iowa Gambling Task. *Neuropsychology*

463   *Review,* 19: 102-114. doi: 10.1007/s11065-009-9083-4.

464 Burton L, Plaff D, Bolt N, Hadjkyriacou D, Silton N, Killgallen C. et al. 2010. Effect of

465   gender and personality on the Conners Continuous Performance Test. *Journal of Clinical*

466   *and Experimental Neuropsychology,* 32: 66-70. doi:10.1080/13803390902806568

467 Caroselli, J.S., Hiscock, M., Scheibel, R. S., & Ingram, F. 2006. The simulated gambling

468   paradigm applied to young adults: An examination of university students' performance.

469   *Applied Neuropsychology,* 13: 203-212.

470 Clark DG, Kar J. 2011. Bias of quantifier scope interpretation is attenuated in normal aging and

471   semantic dementia. *Journal of Neurolinguistics,* 24: 411–9.

472 Cohen AL, Shapiro SK. 2007. Exploring the performance differences on the flicker task and the

473   Conners' Continuous Performance test in adults with ADHD. *Journal of Attention*

474   *Disorders, 11*: 49-63.

475 Conners CK. 2004. Multi Health Systems. Conners' Continuous Performance Test II:

476   Technical guide for software manual. New York: Multi-Health Systems.

477 Conners K, Jeff JL. 1999. ADHD in adults and children: The latest assessment and treatment

478   strategies. Kansas City, MO: Compact Clinicals.

479 Danckert J, Stöttinger E, Quehl N, Anderson B. 2012. Right hemisphere brain

480   damage impairs strategy updating. *Cerebral Cortex*, 22: 2745-2260. doi:

481   10.1093/cercor/bhr351

482 Dikmen SS, Heaton RK, Grant I, Temkin NR. 1999. Test-retest reliability and practice effects of

483   the expanded Halstead-Reitan Neuropsychological Test Battery. *Journal of the*

484   *International Neuropsychological Society,* 5: 346-336.

485 Dougherty DM, Marsh DM, Mathias CW. 2002. Immediate and delayed memory tasks: A

486   computerized behavioral measure of memory, attention, and impulsivity. *Behavior*

487   *Research Methods,* 34: 391-398.

488 Earle-Boyer EA, Serper MR, Davidson M, Harvey PD.1991. Continuous performance tests in

489   schizophrenic patients: Stimulus and medication effects on performance. *Psychiatry*

490   *Research,* 37: 47-56.

491 Fernie G, Tumney RJ. 2006. Some decks are better than others: The effect of reinforcer type and

492   task instructions on learning in the Iowa Gambling Task. *Brain & Cognition,* 60: 94-102.

493 Fox CJ, Mueller ST, Grey HM, Raber J, Piper BJ. 2013. Evaluation of a short-form of the Berg

494   Card Sorting test. *PLoS One,* 8(5): e63885. doi: 10.1371/journal.pone.0063885.

495 González-Giraldo Y, Rojas J, Novoa P, Mueller ST, Piper BJ, Adan A, Forero DA. 2014.

496   Functional polymorphisms in BDNF and COMT genes are associated with objective

497   differences in arithmetical functioning in a sample of young adults. *Neuropsychobiology,*

498   70: 152-157. doi: 10.1159/000366483.

499 Greenberg LM, Waldman ID. 1993. Developmental normative data on the Test of Variables of

500   Attention (T.O.V.A.™). *Journal of Child Psychology & Psychiatry,* 34: 1019-1030.

501 Hawthorne MJ, Weatherford DR, Tochkov K. 2011. Effects of explicit and implicit cognitive

502   factors on the learning patterns in the Iowa Gambling Task. *American Journal of*

503   *Psychological Research,* 7: 64-78.

504 Homack S, Riccio CA. 2006. Conners' Continuous Performance Test (2nd ed.; CCPT-II).

505   *Journal of Attention Disorders, 9*:556-558.

506 Karakas S, Yalm A, Irak M, Erzengin OU. 2002. Digit span changes from puberty to old age

507   under different levels of education. *Developmental Neuropsychology, 22*: 423-453.

508 Keneko H, Yoshikawa T, Nomura K, Ito H, Yamauchi H, Ogura M, Honjo S. 2011.

509   Hemodynamic changes in the prefrontal cortex during Digit span task: A Near-Infrared

510   Spectroscopy study. *Neuropsychobiology,* 63: 59-65. doi: 10.1159/000323446.

511 Kuntsi J, Andreou P, Ma J, Börger NA, van der Meere JJ. 2005. Testing assumptions for

512   endophenotype studies in ADHD: reliability and validity of tasks in a general population

513   sample. *BMC Psychiatry* 5: 40. doi:10.1186/1471-244X-5-40.

514 Lezak MD, Howieson DB, Bigler ED, Tranel D. 2012. Neuropsychological Assessment (5th ed),

515   New York: Oxford.

516 Lipnicki DM, Gunga HC, Belavy DL, Felsenberg D. 2009a. Decision making after 50 days of

517   simulated weightlessness. *Brain Research,* 1280: 84-89. doi:

518   10.1016/j.brainres.2009.05.022.

519 Lipnicki, D. M., Gunga, H. C., Belavý, D. L., & Felsenberg D. (2009b). Bed rest and

520   cognition: effects on executive functioning and reaction time. *Aviation Space &*

521   *Environmental Medicine,* 80: 1018-24.

522 Lynn, R., & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g

523   defined as working memory capacity. *Intelligence,* 36: 226-235.

524 Lyvers M, Tobias-Webb J. 2010. Effects of acute alcohol consumption on executive cognitive

525   functioning in naturalistic settings. *Addictive Behavior,* 35: 1021-28.

526 Mackworth NH. 1948. The breakdown of vigilance during prolonged visual search. *Quarterly*

527   *Journal of Experimental Psychology,* 1: 6-21.

528 McGee RA, Clark SE, Symons DK. 2000. Does the Conners' Continuous performance test aid in

529   ADHD diagnosis? *Journal of Abnormal Child Psychology,* 28: 415-424.

530 Miller DI, Talbot V, Gagnon M, Messier C. 2013. Administration of neuropsychological tests

531   using interactive voice response technology in the elderly: Validation and limitations.

532   *Frontiers in Neurology,* 4: 107. doi: 10.3389/fneur.2013.00107.

533 Mueller ST. 2012. The PEBL Manual, Version 0.13. Lulu Press. ISBN 978-0557658176.

534 Mueller ST. 2015. The Psychology Experiment Building Language, Version 0.14. Retrieved

535   from http://pebl.sourceforge.net.

536 Mueller ST, Piper BJ. 2014. The Psychology Experiment Building Language (PEBL) and PEBL

537   test battery. *Journal of Neuroscience Methods,* 222: 250-259.

538   doi:10.1016/j.jneumeth.2013.10.024.

539 National Science Foundation 2015. Dissemination and sharing of research results. Accessed

540   7/22/2015 at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp

541 Ogg RJ, Zou P, Allen DN, Hutchins SB, Dutkiewicz RM, Mulhern RK. 2008. Neural correlates

542   of a clinical continuous performance test. *Magnetic Resonance Imaging,* 26: 504-512.

543 Paul RH, Lawrence J, Williams LM, Richard CC. 2005. Preliminary validity of "Integneuro™":

544   A new computerized battery of neurocognitive tests. *International Journal of*

545   *Neuroscience,* 115:1549-1567.

546 Peterson DS, Fling BW, Mancini M, Cohen RG, Nutt JG, Horak FB. 2015. Dual-task

547   interference and brain structural connectivity in people with Parkinson's disease who

548   freeze. *Journal of Neurology, Neurosurgery & Psychiatry*, 86: 786-792.

549 Piper BJ. 2010. Age, handedness, and sex contribute to fine motor behavior in children. *Journal*

550   *of Neuroscience Methods,* 195: 88-91. doi: 10.1016/j.jneumeth.2010.11.018.

551 Piper BJ, Acevedo SF, Craytor MJ, Murray PW, Raber J. 2010. The use and validation of the

552   spatial navigation Memory Island test in primary school children. *Behavioural Brain*

553   *Research,* 210: 257-262. doi: 10.1016/j.bbr.2010.02.040.

554   Piper BJ, Acevedo SF, Kolchugina GK, Butler RW, Corbett SM, Honeycutt EB, et al. 2011.

555   Abnormalities in parentally rated executive function in methamphetamine/polysubstance

556   exposed children. *Pharmacology, Biochemistry, & Behavior,* 98: 432-439. doi:

557   10.1016/j.pbb.2011.02.013.

558   Piper BJ, Parkhurst D, Greenhalgh J, Gelety C, Birkett MA. 2015. A neurobehavioral

559   comparison of the Psychological Assessment Resources and the PEBL versions of the

560   Iowa Gambling Task in young-adults. Psychology Experiment Building Language

561   Technical Report, 2015.

562   Piper BJ, Li V, Eowiz M, Kobel Y, Benice T, Chu A, et al. 2012. Executive function on the

563   Psychology Experiment Building Language test battery. *Behavior Research Methods,* 44:

564   110-123. doi: 10.3758/s13428-011-0096-6.

565   Plant RR, Quinlan PT. 2013. Could millisecond timing errors in commonly used equipment be a

566   cause of replication failure in some neuroscience studies? *Cognitive Affective &*

567   *Behavioral Neuroscience 13*: 598-614. doi: 10.3758/s13415-013-0166-6.

568   Richardson JT. 2007. Measures of short-term memory: A historical review. *Cortex,* 43: 635-650.

569   Riccio CA, Reynolds CR, Lowe P, Moore JT. 2002. The continuous performance test: A window

570   on the neural substrates for attention. *Archives of Clinical Neuropsychology,* 17: 235-272.

571   Rosvold HE, Mirsky AF, Sarason I, Bransome ED, Beck LH. 1956. A continuous performance

572   test of brain damage. *Journal of Consulting Psychology,* 20: 343-350.

573   Solanto M, Newcorn J, Vail L, Gilbert S, Ivanov I, Lara R. 2009. Stimulant drug response in the

574   predominantly inattentive and combined subtypes of Attention-Deficit/Hyperactivity

575   Disorder. *Journal of Child and Adolescent Psychopharmacology,* 19: 663–671. doi:

576   10.1089/cap.2009.0033.

577 Soreni N, Crosbie J, Ickowicz A, Schachar R 2009. Stop signal and Conners' continuous

578   performance tasks: test--retest reliability of two inhibition measures in ADHD children.

579   *Journal of Attention Disorders,* 13(2): 137–143. doi: 10.1177/1087054708326110.

580 St Clair-Thompson HL, Allen RJ. 2013. Are forward and backward recall the same? A dual-task

581   study of digit recall. *Memory & Cognition*, 41(4): 519-532. doi: 10.3758/s13421-012-0277-

582   2.

583 Stanislaw H, Todorov N. 1999. Calculation of signal detection theory measures. *Behavior*

584   *Research Methods, Instruments, & Computers,* 31: 137-149.

585 Steingroever H, Wetzels R, Horstman A, Neumann J, Wagenmakers EJ. 2013. Performance of

586   healthy participants on the Iowa Gambling Task. *Psychological Assessment,* 25: 180-193.

587   doi: 10.1037/a0029929.

588 Verdejo-Garcia A, Benbrook A, Funderburk F, David P, Cadet JL, Bolla KI. 2007. The

589   differential relationship between cocaine use and marijuana use on decision-making

590   performance over repeated testing with the Iowa Gambling Task. *Drug & Alcohol*

591   *Dependence,* 90: 2-11.

592 Wardle MC, Hart AB, Palmer AA, de Wit H. 2013. Does COMT genotype influence the effects

593   of d-amphetamine on executive functioning? *Genes, Brain and Behavior,* 12: 13–20. doi:

594   10.1111/gbb.12012.

595 Woods DL, Kishiyama MM, Yund EW, Herron TJ, Edwards B, Poliva O, et al. (2011).

596   Improving digit span assessment of short-term verbal memory. *Journal of Clinical and*

597   *Experimental Neuropsychology,* 33: 101-111. doi: 10.1080/13803395.2010.493149.

599

Figure 1. Scatterplots depicting the association between measures on the Psychology

Experiment Building Language and the Conner's Continuous Performance Test including

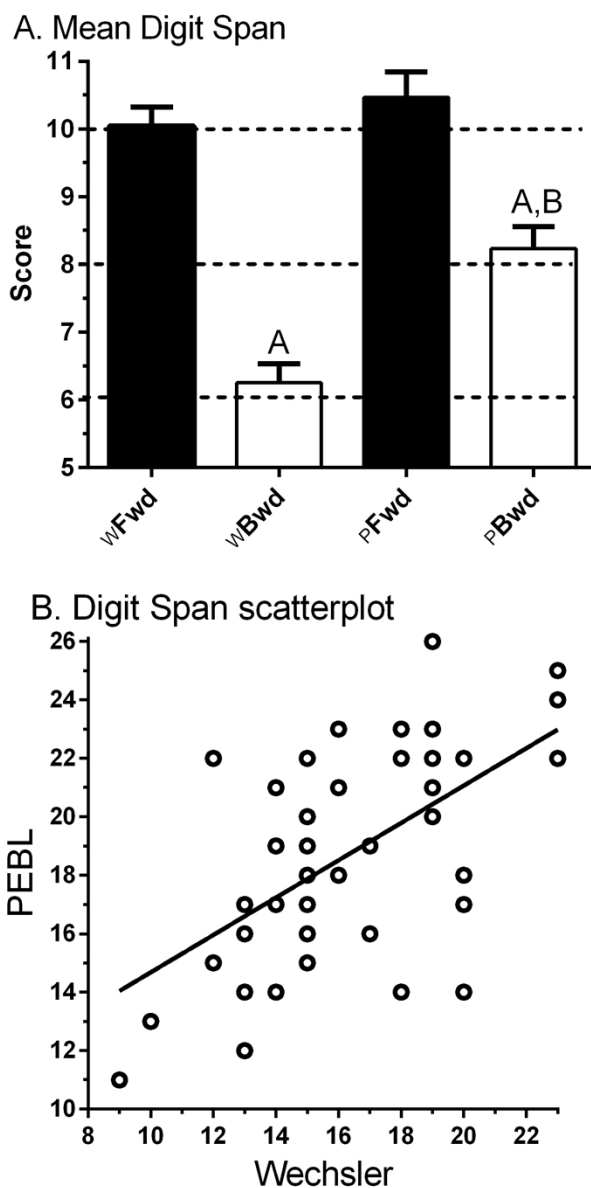reaction time (top), omission errors (middle), and commission errors (bottom).
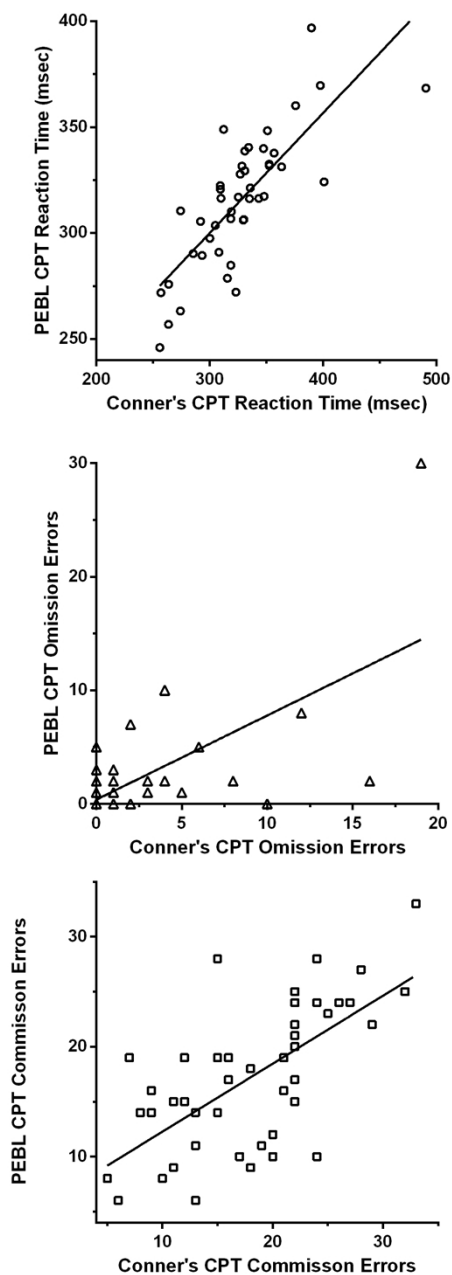
603

604

605   **Figure 2.** A) Wechsler (W) and Psychology Experiment Building Language (P) Digit Span

606   Forward (Fwd) and Backward (Bwd). [A]$p < .0005$ versus Digit Span Forward, [B]$p < .0005$ versus

607   PEBL Digit Span Forward. B) Scatterplot of Wechsler by PEBL Digit Span total ($r_P(45) = .56$, p

608    $< .0005$).

609

611 **Figure 3.** Response times on the Psychological Assessment Resources (PAR, A) and Psychology

612 Experiment Building Language (PEBL, B) Iowa Gambling Task by block of 20 trials ($^*p <$

613 .0005). Selection of advantageous and disadvantageous decks (C, D) ($^*p < .05$ versus

614 disadvantageous on block 1). Selection of each deck (E, F) ($^Ap < .005$ versus Deck B, C, or D;

615 $^Bp < .05$ versus Deck C and D; $^Cp < .05$ versus Deck B). Compensation by trial (G) (horizontal

616 line indicates $p < .05$).

617

619
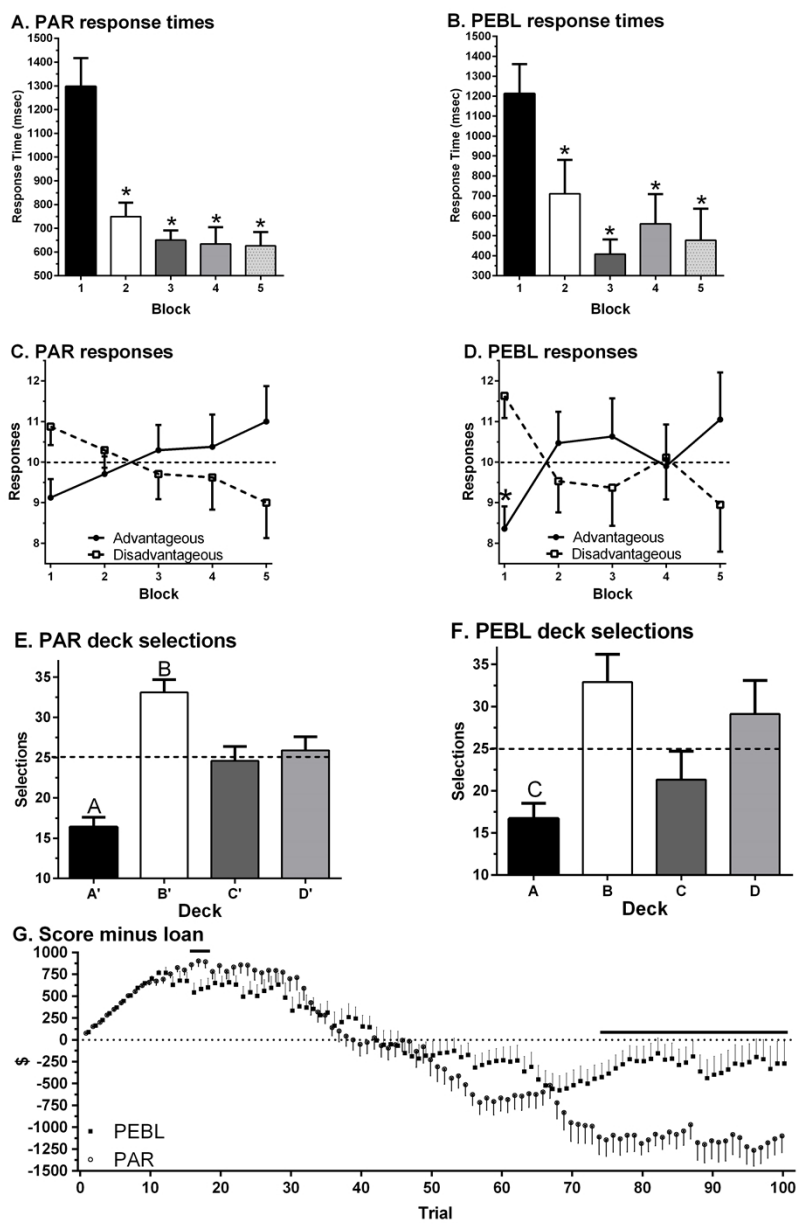
Figure 4. Deck selections over one-hundred trials for the participant (a 34 year-old, Native

American female) with the median NET1 to 5 (0) on the Psychological Assessment Resources

(PAR) Iowa Gambling Task (A). Deck selections for the participant (a 18 year-old Native

American male) with the median NET1 to 5 (-2) on the Psychology Experiment Building

Language (PEBL) Iowa Gambling Task (B).

625
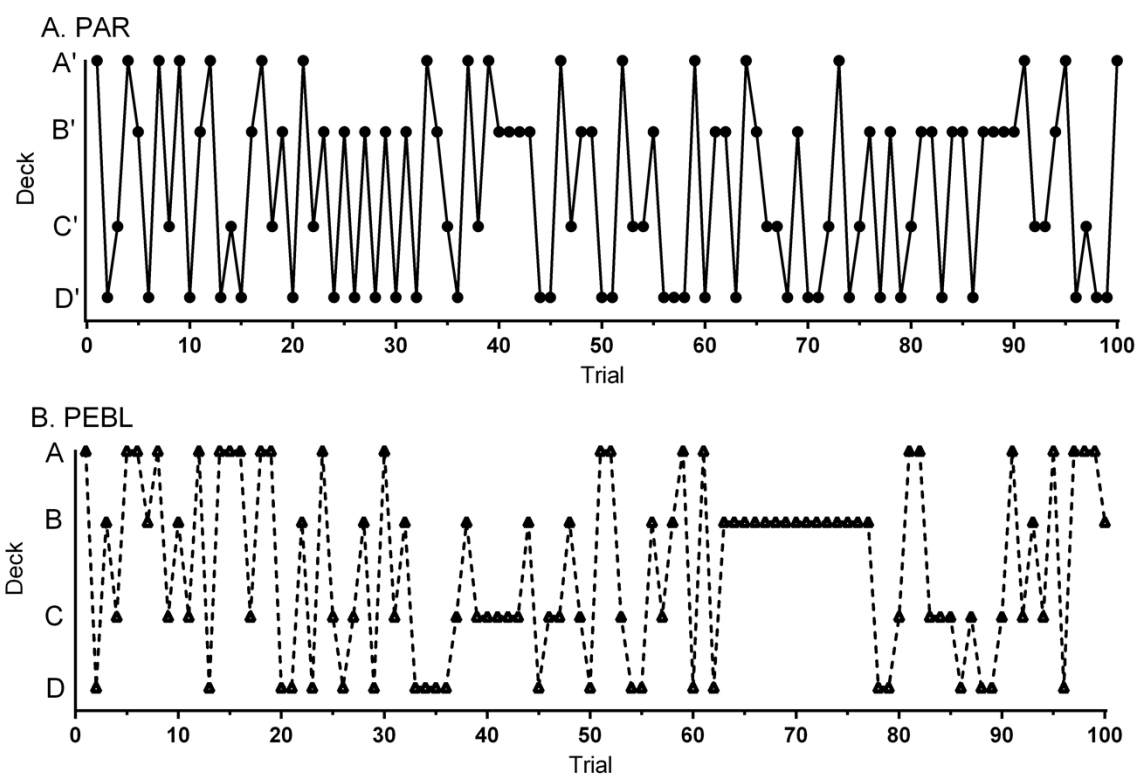
626 **Table 1.** A comparison of the Bechara IGT distributed by Psychological Assessment Resources (PAR) and the Mueller and Bull IGT

627 distributed with version 0.14 of the Psychology Experiment Building Language (PEBL).

628
629

	PAR	PEBL
Instructions (words)	441	379
Visual post-trial feedback	yes	yes
Auditory post-trial feedback	yes	yes
Post-trial wait period	yes	yes
Deck A: Reward ($)	80, 90, 100, 110, 120, 130, 140, 150, 160, 170	100
Deck A: Punishment ($)	150, 200, 250, 300, 350	150, 200, 300, 350
Deck B: Reward ($)	80, 90, 100, 110, 120, 130, 140, 150, 160, 170	100
Deck B: Punishment ($)	1,250, 1,500, 1,750, 2000, 2,250, 2500	1,250
Deck C: Reward ($)	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95	50
Deck C: Punishment ($)	25, 50, 75	25, 50, 75
Deck D: Payoff ($)	40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95	50
Deck D: Loss ($)	250, 275, 300, 350, 275	250
Trials	100	100
Cards/deck (maximum)	60	100
Standardized (T_{50}) scores	yes	no
Cost	$560[P]	$0

657

658   ᴾPrice in U.S.D. on 8/22/2015.

659

660 **Table 2**. Percentiles of the participants (N = 44) on the Conner's Continuous Performance

661 Test. SE: standard error.

662

	<u>Min</u>	<u>Max</u>	<u>Mean</u>	<u>SEM</u>
Reaction time	1.0	94.2	18.6	2.9
Reaction time SE	1.0	99.0	44.3	5.0
Omissions	20.8	99.0	47.5	3.7
Commissions	19.0	99.0	74.4	3.7
d'	10.9	97.3	69.6	3.3
B	24.7	78.1	36.0	1.6

671

672

673 **Table 3**. Intra-test Continuous Performance Test Spearman correlations (Conners/PEBL).

674 [a]$p < .05$.

675

	A.	B.	C.
A. Reaction-Time (msec)	+1.00		
B. Reaction-Time SE	+0.54[a] / +0.18	+1.00	
C. Omission Errors	+0.20 / +0.03	+0.53[a] / +0.35[a]	+1.00
D. Commission Errors	-0.38[a] / -0.36[a]	+0.16 / +0.29	+0.32[a] / +0.36[a]