An updated version of the Weigl discriminates adults with dementia from those with mild impairment and healthy controls

Leigh J. Beglinger a,*, Frederick W. Unverzagtb, Xabier Beristain c, David Kareken c, **

a Department of Psychiatry, University of Iowa College of Medicine, MEB 1-321, Iowa City, IA 52242-1000, USA
b Department of Psychiatry, Indiana University School of Medicine, USA
c Department of Neurology, Indiana University School of Medicine, Neuropsychology Section, 541 Clinical Drive, CL 299, Indianapolis, IN 46202, USA

Accepted 13 December 2006

Abstract

Dementia screening batteries often fall short on measures of executive functioning. The Weigl Color Form Sorting Test (WCFST) is a candidate for inclusion in such batteries, but can be insensitive to mild disturbance. The WCFST consists of 12 colored geometric shapes and requires the patient to sort the pieces by color or form, and then shift to the other sorting principle unassisted. We created a modified version of the WCFST (the Weigl-R) with increased conceptual complexity by adding two stimulus dimensions (texture and central shapes). The range of scores was also increased by adding the extent of examiner assistance required to achieve a correct sort, ability to verbalize conceptual strategy, and number of perseverations. We administered the Weigl-R to a group of 30 patients with mixed dementias, 34 adults with cognitive impairment without dementia, and 21 healthy controls. The new measure discriminated well between healthy controls and older adults with either cognitive impairment without dementia, or dementia. The Weigl-R may be a useful adjunct to brief dementia batteries but requires further validation.

Keywords: Dementia; Executive functioning; Neuropsychology; Conceptual reasoning

Brief assessments of dementia, such as the Consortium to Establish a Registry for Alzheimer Disease battery (CERAD) (Morris et al., 1989), the Repeatable Battery for the Assessment of Neuropsychological Status (Randolph, 1998) and the Blessed Dementia Scale (Blessed, Tomlinson, & Roth, 1968), are widely used and effective instruments for the detection of dementia in older adults. However, measures of conceptual reasoning and mental flexibility are not included in these protocols, despite compromised executive skills in patients with Alzheimer disease and other neurodegenerative disorders such as Parkinson’s disease, Huntington’s disease, vascular dementia and Diffuse Lewy...

** Accepted for publication by Wm. Drew Gouvier.
* Corresponding author. Tel.: +1 319 335 8765; fax: +1 319 353 3003.
** Corresponding author.
E-mail addresses: leigh-beglinger@uiowa.edu (L.J. Beglinger), dkareken@iupui.edu (D. Kareken).
Body disease (Cools, Barker, Sahakian, & Robbins, 2001; Lange, Sahakian, Quinn, Marsden, & Robbins, 1995; Newman et al., 2001; Stout, Rodawalt, & Siemers, 2001). While one exception is the Dementia Rating Scale (Mattis, 1988), the extent to which mental set shifting and conceptual flexibility are measured is limited. Thus, the need remains for additional brief measures of executive functioning that might supplement existing screening batteries, particularly since assessment of executive functioning helps gauge the functional impact of neurodegenerative diseases, the extent of decline, and type of dementia [e.g., conceptual reasoning/set shifting is affected early in subcortical dementias (Cronin-Golomb, Corkin, & Growdon, 1994)]. Predicting conversion to dementia also increases if tasks tap executive functions (Rosas et al., 2001) (Albert, Moss, Tanzi, & Jones, 2001; Amieva et al., 2004; Guarch, Marcos, Salamero, & Blesa, 2004).

The Wisconsin Card Sorting Test (WCST) is a commonly used measure of executive function, but it is time consuming and has floor effects in impaired patients, which are drawbacks for brief dementia evaluations. Shorter forms of the WCST (e.g., WCST-64; Greve, 2001) may also involve more modest reliability and floor effects (Axelrod, 2002; Robinson, Kester, Saykin, Kaplan, & Gur, 1991). The more recently developed Card Sorting Test from the Delis–Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001) is a promising assessment of conceptual reasoning and set shifting (Homack, Lee, & Riccio, 2005; Houston et al., 2005; McDonald, Delis, Norman, Tecoma, & Iragui-Madozi, 2005) although the conceptual demands of this measure can be high. A much simpler alternative more suitable for brief screening batteries is the Weigl Color Form Sorting Test (WCFST; Weigl, 1941) in which 12 colored geometric shapes are first sorted according to one of two dimensions (color or form), and then subsequently sorted to the remaining principle. It can be administered in 5–10 min and is not significantly affected by normal aging (Laiacona, Inzaghi, De Tanti, & Capitani, 2000; Villardita, Cultrera, Cupone, & Mejia, 1985). Byrne, Bucks, and Cuerden (1998) suggested that the original WCFST was insensitive to subtle impairment, in that some patients with dementia could “pass,” perhaps because of very early stage disease or higher premorbid intellect.

The Weigl was modified to increase its sensitivity in 1966 (De Renzi, Faglioni, Savoiardo, & Vignolo, 1966) by expanding the sorting principles from two to five, and including an examiner sort condition for each category, which only required the patient to identify the sorting principle, if they could not sort the chips on their own. This approach resulted in a greater range of scores (0–15), and was effective in the detection of aphasia, but was insensitive to cognitive dysfunction resulting from other causes (Lezak, 1995). An extended scoring system proposed by Grewal & Haward (Grewal, 1984) distinguished demented older adults from controls and depressed patients, but this study did not formally examine gradations of cognitive impairment. However, 12% of depressed patients and 7% of controls scored in the abnormal range, leading the authors to suspect that these subjects had subclinical degenerative disease. Recently, Byrne et al. (1998) explored the long-held view that patients with neurological damage are more likely to sort by color. They modified the Weigl scoring system proposed by Grewal & Haward to include scoring for perseverations to form. They found that a color sorting preference distinguished between dementia patients, amnestic patients, and controls, although sensitivity was low and just over one-third of the moderately demented patients (mean MMSE = 19/30) produced perfect scores on the WCFST with their revised method.

The purpose of the current study was to determine if a newly revised version of the Weigl (Weigl-R) could measure impairments in conceptual reasoning between groups of older adults with varying levels of cognitive impairment. Because of the clinical importance of identifying patients with early, milder symptoms (cognitive impairment/no dementia; CIND), we chose to study patients with both CIND and dementia. To broaden the clinical applicability of our results, we also utilized a wider clinical sample that included mild impairments and dementia of various types, rather than limiting the analyses to those with Mild Cognitive Impairment (MCI) and Alzheimer disease. We hypothesized that this modified version of the Weigl would permit detecting underlying cognitive differences between the three groups, with demented patients performing most poorly, but with CIND patients showing abnormal performance relative to controls.

1. Method

1.1. Participants

Subjects were drawn from the Indiana Alzheimer Disease Center (IADC), a NIA-funded research center. All IADC subjects undergo a comprehensive clinical assessment including physical and neurological examination by a physician, informant interview, neuropsychological testing, and laboratory studies. The neuropsychological assessment
includes the CERAD battery, tests from the Mayo Older Adult Norms battery (Ivnik et al., 1992a,b; Ivnik, Malec, Smith, Tangalos, & Petersen, 1996) and other clinical and experimental cognitive tests, including the Weigl-R. Subjects are diagnosed as normal, cognitive impairment no dementia (CIND) (Graham et al., 1997; Unverzagt et al., 2001) or dementia using a consensus conference format composed of psychiatrists, neurologists, and neuropsychologists. The diagnosis is determined based on a review of the clinical assessment material. Experimental cognitive test data including the Weigl-R are not reviewed in the diagnostic consensus conference. The criteria for dementia are from the DSM-IV (APA, 1994) with dementia subtyped according to presumed etiology using established criteria [e.g., NINCDS/ADRDA for AD (McKhann et al., 1984, NINDS-AIREN for Vascular Dementia (Roman et al., 1993)]. Criteria for CIND are as follows: (a) informant-reported or physician-detected decline in cognition or memory, or (b) psychometric test scores below the approximate 10th percentile of same age peers, and (c) no significant impairment in activities of daily living (i.e., Clinical Dementia Rating equal 0 or 0.5; Morris, 1993). Thus, the CIND group is comprised of a broad array of patients with impairment in at least one cognitive area in the absence of dementia, including patients with MCI, but also those with cognitive impairment from other medical or psychiatric conditions.

Referring diagnoses of the subjects prior to the neuropsychological exam included: Mild Cognitive Impairment (n = 25), probable or possible AD according to NINCDS/ADRDA (n = 28), vascular dementia according to NINDS-AIRENS (n = 4), probable or possible Parkinson’s dementia (n = 8), Progressive Supranuclear Palsy (n = 3), Creutzfeld-Jakob Disease (n = 1), Depression (n = 4), an unknown extrapyramidal syndrome (n = 1), and seizures (n = 1). It is also important to note that referring diagnoses were not always confirmed upon examination and group consensus, thus final sample composition was somewhat different from the referral composition. The Patient sample diagnosed with dementia was significantly less educated than the CIND or control. Therefore, to remove education as a possible confound, those subjects with more extreme levels of education were excluded [n = 15 with <12 years and n = 5 with >19 years of education]. This resulted in a final sample of 30 individuals diagnosed with dementia, 34 with CIND, and 21 healthy controls (Table 1).

### 1.2. Procedure

The Weigl-R has 12 chips in different combinations of colors, shapes, symbols, and textures (Fig. 1). On the first sort, subjects grouped the chips in any manner they chose. After sorting, they were asked to verbalize their sorting strategy. On each of three subsequent trials, they were asked to find an alternate way to sort the chips. If the subject did not provide a correct response after 45 s for any given category (color, shape, symbol, texture), prompts were given as follows: (1) Partial demonstration (examiner completes one group of chips by sorting them to one of the alternate categories—e.g., three red chips) or (2) If the subject could not complete sorting 30 s after the partial demonstration, a direct command was given (e.g., “Please sort the chips by color”). Administration was discontinued if the subject could not correctly sort 30 s after the first direct command.

The score was the sum of sort points (3 = no prompts; 2 = correct to demonstration; 1 = correct to command; 0 = unable) across each of the four categories (total possible = 12). One point was also given for each correct description of the sort (Conceptual Reasoning score, total possible = 4). A perseveration score was also recorded when a previously used category was repeated or the subject reverted to a previously used category during a new sort trial. The perseveration score was not included in the total score.
2. Results

2.1. Subject demographics

There were no significant differences in age between the dementia, CIND, and control groups ($F = 0.61$, ns). As extreme education scores were excluded, there were, by design, no significant group differences for education. The majority of the dementia patients met criteria for the AD subtype (50%). As expected, dementia subjects were significantly more impaired on the MMSE than the CIND and control groups ($F = 58.5$, $p < .0001$) (see Table 1).

2.2. Weigl-R performance in the three groups

Raw scores for the three groups are presented in Table 2. Analysis of variance (ANOVA) using the Weigl-R sorting total score as a dependent variable revealed a significant difference between the three groups ($F[2, 87] = 47.2, p < 0.0001$; Fig. 2), with controls performing better than CIND, who in turn performed better than the dementia participants. All comparisons were significant in post hoc analyses ($p’s < 0.005$). A similar analysis was performed using the Weigl-R Conceptual Reasoning score, which was also significant ($F[2, 87] = 42.8, p < 0.0001$). All post hoc comparisons were again significant ($p’s < 0.005$). The Perseveration score was significantly different between the three groups ($F$

Table 2
Means and standard deviations for Weigl-R Scores by diagnostic group

<table>
<thead>
<tr>
<th>Weigl-R Scores</th>
<th>Controls mean (S.D.)</th>
<th>CIND Mean (S.D.)</th>
<th>Dementia mean (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sorting</td>
<td>11.3 (1.2)</td>
<td>9.3 (2.1)</td>
<td>5.1 (3.3)</td>
</tr>
<tr>
<td>Conceptual reasoning</td>
<td>3.7 (0.45)</td>
<td>3.0 (0.83)</td>
<td>1.5 (1.2)</td>
</tr>
<tr>
<td>Preservation</td>
<td>0.04 (0.21)</td>
<td>0.30 (0.52)</td>
<td>0.53 (0.73)</td>
</tr>
</tbody>
</table>

Note: Total Sorting has 12 possible points; conceptual reasoning has 4 possible points.
Separate ANOVAs of the 4 sorting categories individually showed significant differences between groups \((p = 0.007\) to \(p < 0.0001\)). Post hoc tests of these individual categories showed that texture differentiated controls from CIND \((p = 0.03)\), while color, form and symbol were insignificantly different. Form and symbol differentiated CIND from dementia \((p < 0.0001)\).

2.3. Associations between Weigl-R and other neuropsychological scores

Table 3 shows the inter-relationships between the Weigl-R scores (Sorting Total and Conceptual Reasoning) and other neuropsychological test scores. The patterns of correlations indicate that in CIND, Weigl-R indices do not correlate with other variables, suggesting that the test measures different elements of neuropsychological function. However, in the dementia group, the pattern shows that the highest correlations are with the IU Token Test, a version of

<table>
<thead>
<tr>
<th></th>
<th>MMSE</th>
<th>AF</th>
<th>BNT</th>
<th>CP</th>
<th>WLL1</th>
<th>WLL3</th>
<th>WLLD</th>
<th>Token</th>
<th>AmNART</th>
<th>Sort Total</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE</td>
<td>0.52</td>
<td>0.63</td>
<td>0.40</td>
<td>0.51</td>
<td>0.64</td>
<td>0.44</td>
<td>0.75</td>
<td>−0.43</td>
<td>0.65</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>0.41</td>
<td>0.58</td>
<td>0.41</td>
<td>0.26</td>
<td>0.38</td>
<td>0.30</td>
<td>0.37</td>
<td>−0.30</td>
<td>0.17</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>BNT</td>
<td>0.43</td>
<td>0.38</td>
<td>0.35</td>
<td>0.16</td>
<td>0.41</td>
<td>0.35</td>
<td>0.30</td>
<td>−0.45</td>
<td>0.26</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.16</td>
<td>0.26</td>
<td>−0.07</td>
<td>0.30</td>
<td>0.14</td>
<td>0.06</td>
<td>0.54</td>
<td>−0.28</td>
<td>0.19</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>WLL1</td>
<td>0.43</td>
<td>0.20</td>
<td>0.17</td>
<td>0.27</td>
<td>0.46</td>
<td>0.27</td>
<td>0.42</td>
<td>−0.06</td>
<td>0.26</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>WLL3</td>
<td>0.46</td>
<td>0.37</td>
<td>0.49</td>
<td>0.12</td>
<td>0.60</td>
<td>0.57</td>
<td>0.26</td>
<td>−0.15</td>
<td>0.29</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>WLLD</td>
<td>0.49</td>
<td>0.43</td>
<td>0.47</td>
<td>0.05</td>
<td>0.57</td>
<td>0.72</td>
<td>0.21</td>
<td>−0.16</td>
<td>0.37</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Token</td>
<td>0.41</td>
<td>0.41</td>
<td>0.49</td>
<td>−0.19</td>
<td>−0.09</td>
<td>0.14</td>
<td>0.06</td>
<td>−0.54</td>
<td>0.60</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>NART</td>
<td>−0.54</td>
<td>−0.43</td>
<td>−0.36</td>
<td>−0.20</td>
<td>−0.34</td>
<td>−0.32</td>
<td>−0.36</td>
<td>−0.46</td>
<td>−0.21</td>
<td>−0.21</td>
<td></td>
</tr>
<tr>
<td>Sort total</td>
<td>0.03</td>
<td>0.29</td>
<td>−0.08</td>
<td>0.08</td>
<td>−0.05</td>
<td>0.10</td>
<td>−0.07</td>
<td>0.19</td>
<td>0.01</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Concept</td>
<td>−0.04</td>
<td>0.21</td>
<td>−0.13</td>
<td>0.00</td>
<td>−0.13</td>
<td>0.02</td>
<td>−0.14</td>
<td>0.21</td>
<td>0.12</td>
<td>0.93</td>
<td></td>
</tr>
</tbody>
</table>

Note: Values above the diagonal are patients with dementia; values below the diagonal are patients with CIND. Bold numbers are significant at \(p < 0.05\). MMSE = Mini Mental State Exam; AF = Animal Fluency; BNT = Boston Naming Test; CP = Constructional Praxis; WLL = Word List Learning (Trials 1, 3 and Delay); Token = IU Token Test; AmNART = American National Adult Reading Test.
the Token Test using a printed card rather than separate tokens (see Unverzagt et al., 1999) \( r = 0.60 \), and the MMSE \( r = 0.65 \).

### 3. Discussion

The above findings provide preliminary support for the Weigl-R as a potentially useful means of detecting problems with executive functions in patients with mild cognitive disturbance and dementia, and distinguishing them from healthy older adults. Weigl-R scores were significantly different across all three groups (controls, dementia, and CIND). Weigl-R measures were also unassociated with measures of anterograde learning and memory, language, and constructional praxis in the milder group, supporting the idea that it measures a separate cognitive dimension. Further study of the Weigl-R is nevertheless needed to determine how and to what extent it is associated with performance on other, more common tests of executive function, such as the WCST, parts of the DRS-R, and the Card Sorting Test of the Delis–Kaplan Executive Function battery.

Tests of executive functioning are under-represented in dementia screening batteries, yet executive dysfunction is one of the earliest manifestations of impairment in subcortical dementias, and can increase the predictive power of conversion from MCI to dementia when added to memory tests in patients with Alzheimer’s disease (Stout et al., 2001) (Rosas et al., 2001) (Amieva et al., 2004; Guarch et al., 2004). The Weigl-R may therefore be a useful tool, and has the advantage of brevity over other measures of executive functioning, while also being able to discriminate between mild impairment, dementia and cognitive normality. Also, one of the new stimulus dimensions added in our modified version of the Weigl, texture, was the only dimension to differentiate controls from patients with mild cognitive impairment. Early detection of cognitive impairment is valuable both for patients, who might benefit from early treatment, and for research screening for studies of conversion to dementia. There was clear separation of Weigl-R sort scores between dementia, CIND, and healthy control groups. Although these results need to be replicated in a larger sample, a potential cutoff score of 10 is suggested by this distribution.

Executive functioning is a broad term that includes a variety of components, including planning, organizing, complex problem solving, response set preservation, inhibition, abstraction and reasoning, and initiation and fluency (Lezak, 1995). Past studies have compared the older versions of the WCST to other neuropsychological measures to parse out its specific contributions to executive functioning. When compared with other executive measures (Similarities, Digit Symbol and Trail Making Tests B), only Trails B as a measure of “conceptual flexibility” was found to be a significantly predictor of WCST performance in inpatient adults with alcohol dependence (Tamkin & Dolenz, 1991). In a comparison of the WCST and the Wisconsin Card Sorting Test in healthy adults, Laiacona et al. (2000) concluded that both measures overlap moderately as measures of abstract reasoning, but that the Wisconsin Card Sorting Test has a greater memory load and the WCST requires less sustained attention and greater visuoperceptual skill.

The relatively low intercorrelations of the Weigl-R with measures of memory, praxis, and language suggests that it is tapping unique information. Independent examination of correlations within the dementia group revealed that Weigl-R performance correlated moderately with the MMSE, and with the IU Token Test (language comprehension, working memory). Swihart et al. (1989) found the Token Test to discriminate between healthy adults and patients with Alzheimer’s disease, with performance on this measure correlating highly (0.74) with the MMSE. In our sample, the Token Test and the MMSE were similarly correlated (0.75), the highest correlation in this study.

Some limitations of the current study should be noted. First, as noted above, the Weigl-R is not as established as a measure of executive function, and needs further study in comparison to more established tests of executive skill. Second, the participants in the present sample were a high functioning group with an average of 2 years of college. The generalizability of these findings to other samples of patients with lower levels of education and from different geographic locations is unclear. Our small sample size also precluded a full analysis of sensitivity, specificity, and positive/negative predictive validity. Finally, because this was an exploratory study and we were interested in relationships between the Weigl-R and other neuropsychological measures in the battery, the number of statistical comparisons (Table 3) may have inflated the type I error rate.

In summary, we have shown that the Weigl-R provides a range of scores in patients with varying levels of cognitive function, and that it may provide unique information about cognition. Future research should examine the relationship between the Weigl-R and other more established measures of executive function, its relationship to functional level, and its utility as a predictor of conversion to dementia.
Acknowledgements

The authors would like to thank Elizabeth Way, study coordinator, Nenette Mercado Jessup, technician, and Kathie Fleming, technician, of the Clinical Core of the Indiana Alzheimer Disease Center (IADC). This research was supported by IADC grant P30 AG10133.

References


