Alternate-form reliability of the Dementia Rating Scale-2∗

Kara S. Schmidt a, *, Paul J. Mattis b, Jane Adams c, Paul Nestor c

a Center for Aging, University of Medicine and Dentistry of New Jersey, 42 East Laurel Road, Suite 1500, Stratford, NJ 08084, USA
b Center for Neuropsychological Services, North Shore University Hospital, Manhasset, NY, USA
c Department of Psychology, University of Massachusetts Boston, Boston, MA, USA

Accepted 30 September 2004

Abstract

The Dementia Rating Scale-2 (DRS-2) is a frequently used assessment of cognitive status among older adults in both research and clinical practice. Despite its well-established psychometric properties, its use in serial assessments has posed limitations with regard to practice effects. The primary purpose of the present study is to provide preliminary evidence of alternate-form reliability for the DRS-2. A heterogeneous sample of 52 community-dwelling adults over age 60 with no reported diagnosis of dementia were administered the DRS-2 as well as a newly developed alternate form [DRS-2: AF; Schmidt, K. S. (2004). Dementia Rating Scale-2 Alternate Form: Manual supplement. Lutz, FL: Psychological Assessment Resources]. Our results reveal strong correlations between the two forms; further, no significant differences were found between total scale and subscale scores obtained from the two forms. Therefore, the DRS-2: AF may be a valuable assessment tool in both research and clinical arenas.

Keywords: Dementia; Alzheimer’s disease; Dementia Rating Scale; DRS-2; Alternate-form reliability

© 2004 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

This research was presented at the 31st annual meeting of the International Neuropsychological Society, Honolulu, HI. The data are also presented, in part, in Dementia Rating Scale-2 Alternate Form: Supplemental manual (Schmidt, 2004).

∗ Corresponding author. Tel.: +1 856 566 6843.
E-mail address: schmidka@umdnj.edu (K. S. Schmidt).

0887-6177/$ – see front matter © 2004 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.
The Dementia Rating Scale-2 (DRS-2; Jurica, Leitten, & Mattis, 2001) is a commonly used measure of cognitive status in older adults with neuropsychiatric disease. Over 30 years of research provide support for the instrument’s reliability and validity (e.g., Bobholz & Brandt, 1993; Brown et al., 1999; Chase et al., 1984; Coblentz et al., 1973; Freidl et al., 2002; Gardner, Oliver-Munoz, Fisher, & Emtping, 1981; Green, Woodward, & Green, 1995; Hofer, Piccinnin, & Hershey, 1996; Montgomery & Costa, 1983a; Smith et al., 1994; Vitaliano et al., 1984). The recurrent use of the DRS-2 in serial evaluations, however, may pose limitations with regard to practice effects.

In Meyers et al.’s (1997) study examining the neurotoxic effects of a chemotherapeutic agent over time, scores on the DRS increased over the duration of the study. Total Scores increased from 137.5 (baseline) to 139.7 (after three courses of treatment within a 6-week period). A change of only a few Total Score points can suggest clinically significant improvement on the DRS-2 (Jurica et al., 2001), particularly with higher test scores. Improvements in scores were also observed on the Conceptualization and Memory subscales. Such results are remarkable given the association in the literature between chemotherapy and cognitive impairment (Ahles et al., 2002; Ahles & Saykin, 2002; Anderson-Hanley, Sherman, Riggs, Agocha, & Compas, 2003; Breden, Phillips, Abdolell, Bunston, & Tannock, 2000).

Coblentz et al.’s (1973) research on the test-retest reliability of the original DRS Total Score and subscales in a population of patients diagnosed with Senile Dementia over a 1-week interval further implies the effects of practice. The mean Total Score during the first administration was 79.55; whereas, the mean Total Score after a 1-week retest was 83.18. Increases in scores were also observed on all five subscales. Recent research with patients diagnosed with Alzheimer’s disease (AD) on a semantic fluency task suggests that stability of test performance across serial administrations is more likely with AD patients than the practice effects that occur with controls (Cooper et al., 2001). Therefore, the increase in DRS Total Score is particularly notable in Coblentz et al.’s patients given the severity of dementia in the sample.

Although neuropsychological measures, like the DRS-2, provide essential data with regard to cognitive function at a single point in time, such tools may be limited with respect to their ability to provide accurate documentation of a change in cognition over time due to a lack of alternate or parallel forms (Watson, Pasteur, Healy, & Hughes, 1994). Some attention has been given to the concern of practice effects within the literature (e.g., Basso, Bornstein, & Lang, 1999; Bird, Papadopoulos, Ricciardelli, Rosser, & Cipolotti, 2003; Duff, Westervelt, McCaffrey, & Haase, 2001; Galasko, Abramson, Corey-Bloom, & Thal, 1993; McCaffery et al., 1995; Watson et al., 1994; Wilson, Watson, Baddeley, Emslie, & Evans, 2000); however, insufficient progress has been made in terms of the development of alternate forms for existing neuropsychological measures.

To examine the alternate-form reliability of the DRS-2, the primary goal of this study was to demonstrate equivalence between an alternate form of the DRS-2 (DRS-2: AF; Schmidt, 2004) and the original DRS-2 in a community-dwelling sample of older adults with no reported diagnosis of dementia. It was hypothesized that strong positive correlations would exist between the DRS-2 and DRS-2: AF for Total Score and subscale scores. Likewise, it was hypothesized that no significant differences would exist between the alternate forms on sample means for Total Score and subscale scores.
1. Methods

1.1. Participants

English-speaking individuals over age 60 were recruited from the community through posted flyers and newspaper advertisements in community centers in the greater Boston and Philadelphia areas. All interested participants were screened via a structured telephone interview to meet the following inclusion/exclusion criteria: (1) 60 years of age or over; (2) no history of neuropsychiatric disease, including dementia (e.g., Alzheimer’s type), Huntington’s chorea, Parkinson’s disease, epilepsy, multiple sclerosis, moderate or severe head trauma, schizophrenia or clinically significant depression based on DSM-IV criteria; and (3) no evidence of alcohol-related or drug-dependence disorders based on DSM-IV criteria. A total of 52 (34 females) individuals met the appropriate inclusion/exclusion criteria.

Participants were paid $25 for completing the neuropsychological protocol. The group was composed of 44 Caucasians, 4 African Americans, and 4 Asian Americans. Average age of the sample was 72.8 (S.D. = 7.3), with a range of 60–91. The mean education level was 13.9 years (S.D. = 3.7), with a range of eighth grade to doctorate level. North American Adult Reading Test IQ estimates (Blair & Spreen, 1989) yielded a mean sample Full Scale IQ of 112.2 (S.D. = 8.3), and a range of 96–128. The majority of participants reported a current or past history of caring for a family member with dementia.

1.2. Instruments

The DRS-2 was constructed to assess cognitive abilities in dementia patients, differentiate levels of ability in these patients, and track their cognitive status over time (Jurica et al., 2001). The scale provides a global measure of cognitive functioning as well as five subscale scores: Attention, Initiation/Perseveration, Construction, Conceptualization, and Memory.

The DRS-2: AF (Schmidt, 2004) was designed as an alternate form of the original DRS-2. Thus, the DRS-2: AF includes the same amount of items and the same five subscales that are found in the original form. New questions within the DRS-2: AF reflect the intent of the original questions in the DRS-2. All new word items in the DRS-2: AF verbal recall and verbal recognition sections were based on matching word frequency and syllable length to the original items in the DRS-2 (Kucera & Francis, 1967). Likewise, attempts were made to match all new drawings and figures (e.g., graphomotor designs, construction items, visual matching, visual memory) with the items in the original DRS based on number of straight and curved lines.

1.3. Procedure

Participants were administered a structured clinical interview as well as several brief neuropsychological measures, including the DRS-2 and DRS-2: AF. Both DRS-2 forms were administered in the same session, with approximately 60 min of neuropsychological testing between the two administrations. The two alternate forms were consistently administered in the
same place in the protocol. The investigators counterbalanced which form was administered first to weigh any effects of practice.

To decrease the dropout rate and increase the likelihood of a representative community sample, all participants were tested in their homes rather than in an office or outpatient clinic. Administration and scoring of all tests were completed by a trained graduate student and executed as each respective manual dictates.

2. Results

Table 1 includes the descriptive statistics, including means and standard deviations, for the DRS-2 and DRS-2: AF Total Score and five subscales. The descriptive data across the two alternate is almost identical. Likewise, the Total Score means (standard deviations) for the DRS-2 and DRS-2: AF of 138.20 (5.50) and 137.92 (5.31) are very similar to the original DRS mean of 137.3 (6.9) reported by Montgomery and Costa (1983b) in their sample of controls.

To assess alternate-form reliability, correlations and t tests were run to test for equivalence between the DRS-2 and DRS-2: AF. Pearson product-moment correlations were computed to determine the relationship between the DRS-2 and DRS-2: AF Total Score and the subscale scores. As shown in Table 1, a strong positive relationship exists between the two Total Scores ($r = .82$). The correlations between for the subscale scores were also quite strong. Nonetheless, it should be noted that the Construction subtest was not included in these analyses, or any other analyses, due to its kurtotic distribution. With the exception of two scores, all participants received full credit on the Construction subscale.

Total Score and subscale scores on the DRS-2 and DRS-2: AF were compared using paired t tests to determine whether they produce equivalent scores. The findings revealed no significant differences between the DRS-2 and DRS-2: AF on the Total Score and Attention, Initiation/Perseveration, and Memory subscales. However, there was a slight tendency for the mean scores to differ on the Conceptualization subtest ($t$ ratio $= -1.69$, $p = .097$).

Table 2 shows the percentage of DRS-2: AF scores that fall between 0, 1, 2, 3–5, and 5+ points of the original DRS-2 based on our community-dwelling sample. The greater portion of participants’ scores on the two alternate forms fell within 2 points.

<table>
<thead>
<tr>
<th>Scale/subscale</th>
<th>DRS-2</th>
<th>DRS-2: AF</th>
<th>Alternate form reliability ($r$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ (S.D.)</td>
<td>$M$ (S.D.)</td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>138.08 (5.44)</td>
<td>137.87 (5.25)</td>
<td>.82</td>
</tr>
<tr>
<td>Attention</td>
<td>35.94 (1.11)</td>
<td>36.04 (1.22)</td>
<td>.77</td>
</tr>
<tr>
<td>Initiation/Perseveration</td>
<td>35.90 (1.92)</td>
<td>36.10 (1.55)</td>
<td>.66</td>
</tr>
<tr>
<td>Construction</td>
<td>5.96 (0.19)</td>
<td>5.96 (0.19)</td>
<td>–</td>
</tr>
<tr>
<td>Conceptualization</td>
<td>37.12 (1.95)</td>
<td>36.73 (2.03)</td>
<td>.70</td>
</tr>
<tr>
<td>Memory</td>
<td>23.15 (2.05)</td>
<td>23.04 (2.13)</td>
<td>.80</td>
</tr>
</tbody>
</table>

Note: $N=52$. DRS-2: Dementia Rating Scale-2, AF: Alternate Form. The alternate form reliability correlation coefficient was not computed for Construction due to the subscale’s skewed distribution. All data are raw scores.
Table 2: Percentage of DRS-2: AF scores that fall between 0, 1, 2, 3–5, and greater than 5 points of the original DRS-2

<table>
<thead>
<tr>
<th>Scale/subscale</th>
<th>Points between DRS-2: AF and DRS-2 scores (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Total Score</td>
<td>23.1</td>
</tr>
<tr>
<td>Attention</td>
<td>53.9</td>
</tr>
<tr>
<td>Initiation/Perseveration</td>
<td>67.3</td>
</tr>
<tr>
<td>Construction</td>
<td>96.2</td>
</tr>
<tr>
<td>Conceptualization</td>
<td>30.8</td>
</tr>
<tr>
<td>Memory</td>
<td>53.9</td>
</tr>
</tbody>
</table>

Note: N = 52. DRS-2: Dementia Rating Scale-2, AF: Alternate Form. All data are based on raw scores.

3. Discussion

To our knowledge, the present study is the first to demonstrate alternate-form reliability between the original Mattis Dementia Rating Scale and an alternate form. Our findings suggest that the DRS-2 and DRS-2:AF are strongly correlated in a community-dwelling sample of older adults without a diagnosis of dementia.

Alternate forms of assessment tools, such as the DRS-2, may be valuable within populations of older adults with neuropsychiatric disease. Because there are many explanations for changes in cognitive status (e.g., neurodegenerative disease, infection, metabolic change, depression, medication side effects), proper documentation of symptom progression is imperative for accurate diagnosis and often requires repeated administrations of cognitive status measures. Likewise, individuals presenting with mild cognitive decline may need to be followed over the course of time before a specific dementia diagnosis is given. Progress in the treatment of dementia (e.g., pharmacology, cognitive rehabilitation) and the ongoing research needed to support the efficacy of these treatments necessitates serial administration of cognitive status tools within a relatively brief period of time. Thus, the use of two forms of the DRS-2 may allow for a better classification of cognitive decline and improved evaluation of treatment efficacy.

It should be noted, however, that this study is preliminary in nature. These findings need to be replicated in a sample of patients with documented neuropsychiatric illness (e.g., Alzheimer’s disease, Vascular Dementia). For example, how does the sensitivity and specificity of the DRS-2: AF compare with the documented clinical validity of the original measure (e.g., Green et al., 1995; Vangel & Lichtenburg, 1995)? Additionally, the nature of our study did not permit a random sampling of individuals: a rather large proportion of our sample was relatively well-educated and Caucasian.

Further, our participants were identified as “community-dwelling adults” with “no reported diagnosis of dementia.” This does not imply a non-demented sample, as a participant could meet criteria for dementia without yet having a clinical diagnosis. It is also possible that some subjects misreported their medical history. In addition, many of the participants received relatively “high” scores on the measures, as the distribution of DRS-2 and DRS-2: AF scores tends to be negatively skewed. It is very possible that a greater variability in scores between the two alternate forms would occur in a patient sample.
Lastly, the availability of alternate forms does not necessarily eliminate practice effects altogether. It is likely that some patients who are given serial neuropsychological assessments improve their performance over time because they simply learn how to take the tests; rather than remembering the items over time, patients may learn specific test taking skills. Access to equivalent or alternate forms, however, is apt to reduce improvements in scores due to repeated exposure (e.g., Frazier, Adams, Strauss, & Redline, 2001).

Acknowledgements

The authors thank Dr. Steven Mattis for his helpful comments in regard to the development of the alternate form. The authors also thank Psychological Assessment Resources, Inc. for donating testing supplies for this research project.

References


