Stability of the Jebsen-Taylor Hand Function Test Across Three Test Sessions

Erica B. Stern

Key Words: assessment process, occupational therapy • dexterity functions • reliability of tests • tests, by title, Jebsen-Taylor Hand Function Test

The Jebsen-Taylor Hand Function Test (Jebsen, Taylor, Trieschmann, Trotter, & Howard, 1969) was designed as a broad measure of hand function and is widely used by occupational therapists. The test measures the time taken to perform seven hand tasks. Its test-retest reliability for the dominant hand was reported as ranging from $r = .67$ to $r = .99$ across the seven subtests (Jebsen et al., 1969). As the body of research on and using the Jebsen-Taylor test has grown (Carlson & Trombly, 1983; Lynch & Bridle, 1989; Noronha, Bundy, & Groll, 1989; Rider & Linden, 1988; Spaulding, McPherson, Strauchota, Kuphal, & Ramponi, 1988), questions and issues surrounding test reliability and validity have become more important. In an earlier paper (Stern, 1991), I reported a study in which the Jebsen-Taylor test was used to compare hand function of nondysfunctional subjects when wearing each of four styles of a wrist extensor orthosis and when using an unencumbered hand. The Jebsen-Taylor test data were gathered across three sessions to consider changes that could occur as subjects became used to the orthoses. The primary question of the study was addressed with the use of an analysis of variance, which indicated that the five orthotic conditions afforded significantly different speeds across the seven subtests as well as when total mean test scores were compared (see Table 1).

In an unexpected secondary finding, the seven Jebsen-Taylor subtests were shown to have significantly different patterns of scores across the three sessions ($p = .02$) (see Table 2). This difference could not be ascribed to differing speeds among the orthotic conditions and thus represented a question of test stability across sessions. Because questions of test stability were not central to the original hypotheses, the finding was reported but not fully analyzed in the original article (Stern, 1991). The present paper is intended to analyze the serendipitous finding and discuss its clinical ramifications.

Method

Twenty right-hand-dominant women ranging in age from 18 to 33 years were assessed under five orthotic conditions: (a) free handed, (b) while wearing a Futuro® orthosis, (c) while wearing a volar thumbhole orthosis, (d) while wearing a volar connector bar, and (e) while wearing a dorsal wrist extensor orthosis. Data were gathered on only the dominant hand. The Jebsen-Taylor subtests (i.e., writing, turning cards, moving small objects, simulated feeding, stacking checkers, moving empty cans, and moving 1-lb cans) were administered for each orthotic condition, with data gathered across three test sessions. Test sessions were separated by a minimum of 4 days, and

---

1Manufactured by Kendall-Futuro Company, Suite 900, One Riverfront Placie, Newport, KY 41071–4544.
Total dexterity showed significant improvement (p = .0001) across the three sessions (see Table 1). A post hoc analysis with the Newman-Keuls Multiple Comparisons Test (Winer, 1971) showed each successive session to be significantly faster than its predecessor (p < .01).

The subtests of the Jebsen-Taylor test showed significantly different practice effects (p = .02), though each subtest's third session time was faster than the second, and the second was faster than the first. Mean differences between first and third sessions ranged from .48 sec to .97 sec, with third sessions being 18% faster than first sessions (see Table 2). A post hoc analysis with the Newman-Keuls test showed that differences among sessions reached the p < .05 level of significance for two subtests (writing and simulated feeding). In both of these subtests, the third session times were significantly faster than the first, though not significantly faster than the second.

**Discussion**

Jebsen et al. (1969) reported no significant difference in subtest scores between first and second test sessions when administering the Jebsen-Taylor test to 26 patients with stable hand disorders. The present study affirms that most of the Jebsen-Taylor subtests have strong test-retest reliability. The evidence, however, indicates that the times for the writing and simulated feeding subtests are more questionable across several sessions of testing. The differing results may be related to differences in the subjects' subjects (i.e., my subjects were without hand injuries, whereas Jebsen et al. used subjects with hand injuries) or the number of times that subjects were assessed, or it may reflect a true practice effect for those two subtests. If the instability is a true practice effect, therapists may wish to look more questioningly at patients' improvement in the writing and simulated feeding subtests when not accompanied by similar change in the other subtests.

One must also consider what commonality exists between writing and simulated feeding tasks that could cause them to demonstrate larger practice effects than the other hand tasks. Writing and simulated feeding are the only Jebsen-Taylor subtests that require the subjects to use a tool (i.e., a pen and a spoon). Perhaps nondysfunctional subjects draw from their large repertoire of grasp and pinch styles, refining their approach over several sessions to produce significantly faster times. Jebsen et al.'s (1969) original test-retest reliability for subtests was based on persons with stable hand disorders. That population may have demonstrated less significant practice effects in these subtests because such disorders limit the subjects' grasp and pinch options, thus leaving them less able to adapt their tool handling across test sessions. The possibility that grip and pinch were varied across sessions is supported by an earlier observation that during the simulated feeding subtest, many subjects grasped the spoon using a primitive palmar grip rather than the more typical modified three-point pad pinch associated with spoon use in feeding (Stern, 1991).

**Study Limitations**

Data in this study were gathered on subjects with neither motor nor sensory deficits. Perhaps the practice effects observed with these subjects would not be seen with subjects with disabilities. Additionally, reliability is typically established through a comparison of results across two

---

**Table 1**

Hand Function Analysis of Variance With Three Levels of Repeated Measures

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sessions</td>
<td>165.44</td>
<td>2</td>
<td>82.72</td>
<td>70.20</td>
<td>.0001</td>
</tr>
<tr>
<td>Error</td>
<td>44.78</td>
<td>38</td>
<td>1.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtestsa</td>
<td>11,376.56</td>
<td>6</td>
<td>1,894.09</td>
<td>337.70</td>
<td>.01</td>
</tr>
<tr>
<td>Error</td>
<td>640.07</td>
<td>114</td>
<td>5.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthoses</td>
<td>91.08</td>
<td>4</td>
<td>22.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>51.56</td>
<td>76</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sessions x Subtest 13.72 12 1.14 2.07 .02
Sessions x Orthoses 9.73 8 1.22
Subtests x Orthoses 44.96 24 1.87 4.23 .01
Subtests x Orthoses 19.42 48 0.40 0.87 .73


---

**Table 2**

Subjects' Mean Time (Standard Deviation) in Seconds Across Sessions on Seven Jebsen-Taylor Subtests

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Session</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing</td>
<td>10.03 (.65)</td>
<td>9.37 (.37)</td>
<td>9.06 (.37)</td>
<td></td>
</tr>
<tr>
<td>Turning cards</td>
<td>4.05 (.32)</td>
<td>3.64 (.21)</td>
<td>3.53 (.23)</td>
<td></td>
</tr>
<tr>
<td>Mowing small objects</td>
<td>6.28 (.49)</td>
<td>5.78 (.35)</td>
<td>5.61 (.26)</td>
<td></td>
</tr>
<tr>
<td>Simulated feeding</td>
<td>7.24 (.55)</td>
<td>6.68 (.18)</td>
<td>6.50 (.47)</td>
<td></td>
</tr>
<tr>
<td>Stacking checkers</td>
<td>3.57 (.28)</td>
<td>3.10 (.43)</td>
<td>2.94 (.14)</td>
<td></td>
</tr>
<tr>
<td>Mowing empty cans</td>
<td>3.10 (.15)</td>
<td>2.79 (.10)</td>
<td>2.62 (.80)</td>
<td></td>
</tr>
<tr>
<td>Mowing full cans</td>
<td>3.16 (.13)</td>
<td>2.83 (.07)</td>
<td>2.66 (.08)</td>
<td></td>
</tr>
</tbody>
</table>

test sessions. This study considered repeated testing, thereby assessing the Jebsen-Taylor test with greater rigor than most clinical assessment tools.

Conclusion

This study reinforces Jebsen et al.’s (1969) report of strong test stability for five of the seven Jebsen-Taylor subtests. However, two subtests (writing and simulated feeding) demonstrated less stable results across multiple sessions. Clinicians using the test may wish to view small increases in speed in those two subtests more conservatively, recognizing that such increases could represent variability in the subtest rather than a clear improvement in the patient’s ability. This conservatism appears especially appropriate when improvements are noted exclusively in the writing and simulated feeding subtests. Additionally, persons developing hand function tests based on the Jebsen-Taylor test may wish to consider the writing and simulated feeding subtests as weaker elements within the tool and substitute these tasks with more reliable behaviors.

Acknowledgments

Thanks to Rondell Berkeland, MPH, OTR, Judith Reisman, PhD, OTR, and Virgil Mathiowetz, PhD, OTR, for their support and editorial input. I also thank the subjects and data gatherers who participated in the original study.

References


