**BRIEF REPORT**

Test–Retest Reliability of the Purdue Pegboard for Persons With Multiple Sclerosis

Jennifer Gallus, MS, OTR/L, is Occupational Therapist, Fairview Hospitals, Minneapolis, Minnesota. (Mailing address: 5859 Blaisdell Avenue, Minneapolis, Minnesota 55419; jengallus@excite.com)

Virgil Mathiowetz, PhD, OTR/L, FAOTA, is Associate Professor, Program in Occupational Therapy, University of Minnesota, Minneapolis, Minnesota.

KEY WORDS  
• assessments  
• dexterity tests  
• fine motor dexterity

OBJECTIVE. The Purdue Pegboard test often is used in clinical settings to evaluate changes in clients' fine motor dexterity. The purpose of this study was to determine the test–retest reliability and practice effects of the Purdue Pegboard for persons with multiple sclerosis. In addition, this study compared the reliability of one-trial administration to three-trial administration of the four subtests.

METHOD. Thirty-two volunteers from a midwestern community-based maintenance rehabilitation center for persons with multiple sclerosis participated in this study. The participants were administered the four subtests of the Purdue Pegboard, three trials in a row. A second administration was completed 1 week later. Data from 25 participants were analyzed using paired *t* tests, Pearson product-moment correlations, and intraclass correlation coefficients.

RESULTS. The test–retest reliability coefficients ranged from .85 to .90 for one-trial administration and from .92 to .96 for the sum of three trials. No significant practice effects existed except for the sum of three trials of both hands.

CONCLUSION. This study suggests that the one-trial administration of the Purdue Pegboard is a sufficiently reliable assessment to use with persons with multiple sclerosis. Findings further suggest that for a person with multiple sclerosis, any changes in Purdue Pegboard scores using one-trial administration may reflect actual change in that person's dexterity, as no practice effect was demonstrated in this study.


Many occupational therapists use standardized dexterity tests to evaluate the fine or gross motor abilities of their clients. The majority of the research on reliability and validity of these tests has involved persons without disabilities. However, the tests often are used for persons with disabilities for whom they were not standardized. It cannot be assumed that the reliability, validity, and possible practice effects of assessments are the same for persons with and without disabilities (Portney & Watkins, 2000).

Persons with multiple sclerosis may be referred to an occupational therapist because of problems in upper-extremity motor performance that may affect their functional abilities. The Purdue Pegboard, an assessment used to measure upper-extremity fine motor dexterity as well as gross motor coordination (Tiffin, 1968), may be used with this population. No research has evaluated the reliability and validity of this test for persons with multiple sclerosis.

Test–Retest Reliability of the Purdue Pegboard

Historically, test–retest reliability of assessments has been evaluated using correlation coefficients to assess the relationship or consistency between two test occasions. If the researcher was concerned about a practice or learning effect (i.e., individuals scored better on an assessment simply because they had taken it before), they also used *t* tests to assess the average agreement between two test occasions. In recent years, the intraclass correlation coefficient (ICC) has become more popular because it evaluates both aspects of test–retest relia-
bility using one index (Portney & Watkins, 2000).

The test–retest reliability of the Purdue Pegboard has been established as being lower for persons without disabilities (Bass & Stucki, 1951; Buddenberg & Davis, 2000; Desrosiers, Bravo, & Dutil, 1995; Reddon, Gill, Gauk, & Maerz, 1988; Tiffin, 1968; Tiffin & Asher, 1948) than for persons with disabilities, such as rheumatoid arthritis and mental retardation (Guarnaccia, Daniels, & Sefick, 1975; Jones et al., 1991), when using a one-trial administration. In addition, Tiffin and Asher (1948) reported higher test–retest reliability on the Purdue Pegboard using three-trial administration (.82–.91) than a one-trial administration (.60–.76) for persons without disabilities. However, the three-trial administration reliability scores were calculated using Spearman-Brown statistics, not by actually having the participants complete three trials. Recently, Buddenberg and Davis (2000) reported higher test–retest reliability using a three-trial administration (ICC = .81–.89) than a one-trial administration (ICC = .37–.61) for persons without disabilities. These results indicate that three-trial administration should be used with persons without disabilities. Therapists need to know whether three-trial administration is necessary to use for persons with disabilities, such as multiple sclerosis.

Finally, therapists who use the Purdue Pegboard as an assessment need to know whether a practice effect exists before they assume improvement is due to a therapy intervention. If a practice effect exists, improvement in the score may be because individuals had completed the assessment previously and not because their dexterity had actually improved. No study has evaluated the practice effects of the Purdue Pegboard for persons with multiple sclerosis.

Purpose

The purpose of this study was to evaluate the test–retest reliability and practice effects of the Purdue Pegboard with one-trial and three-trial administrations for persons with multiple sclerosis. The results of this study are intended to determine whether the Purdue Pegboard is reliable as a measure of dexterity and whether three-trial administration is needed to achieve acceptable reliability for this population.

Method

Participants

Thirty-two persons with progressive multiple sclerosis were recruited from a maintenance rehabilitation center for persons with multiple sclerosis in the midwestern United States. To attend the center, persons have to score between 5 and 8 on the Kurtzke Expanded Disability Status Scale (Di Fabio, Soderberg, Choi, Hansen, & Schapiro, 1998; Kurtzke, 1983), which means that they have moderate to severe disability.

To be included in the study, participants needed to have multiple sclerosis as diagnosed medically and have sufficient dexterity to complete one peg placement on each of the subtests of the Purdue Pegboard (one-trial and sum of three trials). The data from 7 participants were excluded from analysis: 5 participants did not have sufficient dexterity to place one peg; 1 was excluded for being unable to complete the test using the standard set-up; and 1 was not available for retesting. Twenty-five participants (3 men, 22 women), ranging in age from 30 to 69 years (M = 55.32 years), were included in the data analysis. Twenty-three participants were right-hand dominant, and 2 were left-hand dominant. The time since the participants received a diagnosis of multiple sclerosis ranged from 3 to 25 years (M = 15.72 years). Participants reported that they had no previous upper-extremity injury that affected their function, and they had no exacerbation of symptoms during the course of the study.

Instrument

The Purdue Pegboard used was a model 32020 manufactured by Lafayette Instrument Company (Tiffin, 1968). The extreme right and left side cups each contained 25 pegs. The cup second from the left held 40 washers, and the cup second from the right held 20 collars for the right-handed participants and vice versa for the left-handed participants. The unimanual or one-handed tasks of the Purdue Pegboard consist of placing as many pegs as possible in the column corresponding to the hand being tested. The dominant hand is tested first followed by the nondominant hand. In the bimanual (both hands) subtest, the dominant and nondominant hands are used simultaneously to place pegs in both columns. The assembly subtest involves picking up and placing pegs, washers, and collars using alternating hands. The person has 30 sec to complete each of the first three subtests: the dominant hand, nondominant hand, and bimanual. The scores on these tests are the number of pegs, or pairs of pegs for the bimanual subtest, placed within the time limit. The assembly subtest is limited to 60 sec. This score represents the number of pieces assembled (i.e., pin, washer, collar, second washer). A fifth subtest was not analyzed in this study. It is not another actual test administration but, rather, the sum of the dominant, nondominant, and bimanual scores.

Procedure

The participants were seated at a table with the Purdue Pegboard directly in front of them. The evaluator was seated to their right. The evaluator explained the purpose and procedure of the study, and consent was obtained. The hand used to sign the consent form identified the participants’ dominant hand. The standardized procedure for administering the Purdue Pegboard (Tiffin, 1968) was followed. Participants were given the opportunity to practice each subtest before the timed test to ensure understanding. Each subtest was administered three times in a row. The first trial of the three from each subtest was analyzed as the one-trial score.

The participants were tested during 1 week and retested 1 week later at the same time of day. One week between test and retest was chosen as being the shortest time that a therapist might choose to retest a client. This 1-week period is short enough to reduce the likelihood of an exacerbation of the multiple sclerosis symptoms, while being long enough to minimize possible practice effects. Because multiple sclerosis is a disease of remission and exacerbation, any time span longer than 1 week could result in changes due to the disease process. The same evaluator, Purdue Pegboard apparatus, and stopwatch were used for all test administrations.
Data Analysis

Test–retest reliability. The participants’ one-trial and sum-of-three-trials scores for each of the four Purdue Pegboard subtests were analyzed using Pearson product-moment correlations ($r$) and ICCs. Pearson correlation coefficients were used to determine whether a relationship exists between the test and retest scores (Portney & Watkins, 2000) and to allow for easy comparison with previous test–retest reliability research (e.g., Reddon et al., 1988). In addition, the data were analyzed using the ICC (3,1) because it accounts for both the relationship and the agreement between scores and is a more comprehensive measure of reliability (Portney & Watkins, 2000). Test–retest reliability coefficients ($r$, ICC) less than .50 represent poor test–retest reliability; .50 to .75 indicates moderate test–retest reliability; and .75 to 1.00 demonstrates good test–retest reliability. A score of .90 is considered ideal (Portney & Watkins, 2000).

Practice effect. Data were analyzed to determine whether a practice, or learning, effect exists when the Purdue Pegboard is used to test persons with multiple sclerosis. A paired $t$ test was run on test–retest data for the one-trial and sum-of-three-trials scores of each subtest. The level of significance was set at $p < .05$.

Results

Descriptive data on test and retest scores using one-trial and sum-of-three-trials scores are presented in Table 1. For one-trial administration, the $r$ and ICC values ranged from .85 to .90. For the sum of three trials, the range of $r$ and ICC values was .92 to .96. For all four subtests, the sum-of-three-trials correlations were somewhat higher than the one-trial correlations. The paired $t$ tests between test and retest scores were not significant except for the sum of three trials for the bimanual subtest, indicating a practice effect for this subtest only.

Discussion

Test–Retest Reliability

The test–retest reliability of the Purdue Pegboard using one-trial administration has been reported as ranging from .37 to .92 in healthy populations (Bass & Stick, 1951; Buddenberg & Davis, 2000; Desrosiers et al., 1995; Reddon et al., 1988; Tiffin, 1968; Tiffin & Asher, 1948; Wilson, Iacoviello, Wilson, & Risucci, 1982). .73 to .87 in persons with rheumatoid arthritis (Jones et al., 1991), and .71 to .96 in persons with mental retardation (Guarnaccia et al., 1975). The test–retest reliability of the Purdue Pegboard using three-trial administration has been reported as ranging from .82 to .89 in healthy persons (Buddenberg & Davis, 2000). The results of this study demonstrate that the test–retest reliability of the Purdue Pegboard when used for persons with multiple sclerosis is higher than when used with healthy persons, persons with rheumatoid arthritis, and persons with mental retardation.

The subtests administered in this study demonstrated test–retest reliability between $r = .849$ to .961 and ICC = .847 to .961, which reflects good to excellent reliability. One-trial administration of all the subtests typically takes about 10 min total time, whereas three-trial administration of all the subtests takes about 20 min total time. For most clinical purposes, where treatment time is limited, the test–retest reliability of one-trial administration appears to be adequate. For research purposes, where the highest possible test–retest reliability is important, the sum-of-three-trials administration appears to be more reliable.

Practice Effect

Previous studies involving the Purdue Pegboard have shown no consistent pattern of a practice effect. For persons without disabilities, Tiffin and Asher (1948) reported no practice effect; Jones et al. (1991) reported a practice effect for the dominant hand only; Reddon et al. (1988) reported a practice effect for the dominant and nondominant subtests with men and for the assembly subtest with women; and Wilson et al. (1982) reported a practice effect for the bimanual task with children. For persons with disabilities, Stern, Ytterberg, Krug, and Mahowald (1996) reported a practice effect for the assembly subtest only for persons with rheumatoid arthritis.

In the present study, the only subtest of the Purdue Pegboard that indicated a significant practice effect between the two administrations was the sum of three trials for the bimanual subtest ($t = -2.85$). In contrast, one subtest (nondominant hand, one-trial) showed a nonsignificant decrease in the mean scores between test and retest ($t = .40$). Thus, no consistent trend of a

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Initial Test $M$ (SD)</th>
<th>Retest $M$ (SD)</th>
<th>Difference $M$ (SD)</th>
<th>$t$</th>
<th>$r$</th>
<th>ICC (3,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant hand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One trial</td>
<td>7.32 (3.00)</td>
<td>7.44 (3.23)</td>
<td>-0.12 (1.62)</td>
<td>-0.371</td>
<td>.868</td>
<td>.865</td>
</tr>
<tr>
<td>Sum of three trials</td>
<td>22.60 (9.70)</td>
<td>22.64 (9.81)</td>
<td>-0.04 (3.68)</td>
<td>-0.054</td>
<td>.929</td>
<td>.929</td>
</tr>
<tr>
<td>Nondominant hand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One trial</td>
<td>5.84 (2.76)</td>
<td>5.72 (2.59)</td>
<td>0.12 (1.48)</td>
<td>0.405</td>
<td>.849</td>
<td>.847</td>
</tr>
<tr>
<td>Sum of three trials</td>
<td>18.32 (8.75)</td>
<td>18.33 (8.11)</td>
<td>-0.04 (3.36)</td>
<td>-0.060</td>
<td>.923</td>
<td>.921</td>
</tr>
<tr>
<td>Bimanual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One trial</td>
<td>4.20 (2.31)</td>
<td>4.64 (2.63)</td>
<td>-0.44 (1.36)</td>
<td>-1.622</td>
<td>.857</td>
<td>.850</td>
</tr>
<tr>
<td>Sum of three trials</td>
<td>12.84 (7.28)</td>
<td>14.00 (7.33)</td>
<td>-1.16 (2.03)</td>
<td>-2.851*</td>
<td>.961</td>
<td>.961</td>
</tr>
<tr>
<td>Assembly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One trial</td>
<td>12.12 (6.98)</td>
<td>12.12 (7.22)</td>
<td>0.00 (3.11)</td>
<td>0.000</td>
<td>.905</td>
<td>.904</td>
</tr>
<tr>
<td>Sum of three trials</td>
<td>36.36 (22.29)</td>
<td>38.12 (23.75)</td>
<td>-1.76 (8.26)</td>
<td>-1.066</td>
<td>.938</td>
<td>.935</td>
</tr>
</tbody>
</table>

Note. $N = 25$.

*p < .01

January/February 2003, Volume 57, Number 1
practice effect was found across all subtests and trials. Therefore, this study suggests that any change in the Purdue Pegboard score when used with persons with multiple sclerosis may be attributed to actual changes in the person and not to a practice effect, with the exception of sum of three trials for the bimanual subtest.

Limitations

Several limitations exist for this study. Because of the admission criteria of the center from which the participants were recruited, this sample of convenience may not be representative of all persons with multiple sclerosis. The participants had progressive multiple sclerosis with resulting moderate to severe disability. The results from this study, therefore, may not be applied to persons with mild or nonprogressive multiple sclerosis. There was a higher percentage of women in the sample (88%) than in the general population of persons with multiple sclerosis (60%) (Hietpas, Hooks, Atchison, Pedretti, & McCormack, 1996). The small sample size may also have influenced the results. Portney and Watkins (2000) stated that a sample size of 30 typically is considered acceptable.

A variety of variables could have affected the participants’ Purdue Pegboard scores across testing times, including exacerbations and remissions of symptoms not reported by the participants, fatigue level, room temperature, and events that occurred before the testing. The 1-week interval minimized the effect that exacerbations or remissions may have had on the results, but a change in performance was still possible. However, the high test–retest reliability and the lack of an overall practice effect suggest that these variables had a minimal effect on the results.

The participants in this study displayed a wide range of dexterity ability (see large standard deviations in Table 1). The heterogeneity of the research population could have inflated the test–retest reliability (Portney & Watkins, 2000).

Directions for Future Research

The reliability of the Purdue Pegboard for assessing persons with disabilities other than multiple sclerosis needs to be examined. The relationship between a person’s Purdue Pegboard score and functional performance needs further research. As payment for occupational therapy services changes, research must document that an improvement in dexterity is related to an improvement in functional performance.

Conclusion

This study suggests that the one-trial administration of the Purdue Pegboard is a sufficiently reliable assessment to use with persons with multiple sclerosis. The findings further suggest that changes in Purdue Pegboard scores for a person with multiple sclerosis using one-trial administration may reflect actual change in that person’s dexterity, as no practice effect was demonstrated in this study. ▲

Acknowledgments

We thank Peggy Mueller, OTR/L, and the rest of the staff at the Multiple Sclerosis Achievement Center, St. Paul, Minnesota. We especially thank all the persons who participated in this study.

This paper was written in partial fulfillment for a master of science degree in occupational therapy at the University of Minnesota, Minneapolis.

References


The American Journal of Occupational Therapy

111