Test-Retest Reliability
Study of the
Pennsylvania Bi-Manual Worksample

Stanley Paul

Key Words: child development disorders • work evaluation

Test-retest reliability of a vocational assessment battery called the Pennsylvania Bi-Manual Worksample (PBWS) (Roberts, 1969) was examined in this study. The PBWS was administered to 30 adults with developmental disabilities at two different times, with an interval of 7 to 10 days between administrations. The subjects, who had mild to moderate degrees of mental retardation and ranged in age from 25 to 51 years, were chosen randomly from a group of 95 employees of a sheltered workshop for adults with developmental disabilities in western New York. When the scores from the two test administrations were compared to determine test-retest reliability, a high intraclass correlation value of .98 was found, which indicates that the PBWS provides consistent information about the vocational skills of adults with developmental disabilities.

Habilitation focuses on the intrinsic worth and dignity of the person. According to Jaques (1970), habilitation refers to an initial educational learning process for persons born with a disability or persons who have acquired a disability early in life. Habilitation is usually defined as a process by which professionals help persons with disabilities to maximize their capacities (Rosen, Clark, & Kivitz, 1977) so they can function more effectively as family members and members of society.

Work Evaluation and Occupational Therapy

Much of an adult’s time is spent on work or work-related activities, including social activities that often originate in the work environment. Employment is crucial for adults with developmental disabilities, because they may believe that they are not true adults without it (Schalock & Kiernan, 1990; Turner, 1987). Persons with developmental disabilities are often perceived as lacking the necessary work skills to obtain jobs. Many of them have these skills, which must be identified through evaluations. Because identification and evaluation of vocational skills are necessary steps for job obtaining (Cardner & Chapman, 1990), work potential evaluation instruments have become part of the habilitation process of many persons with disabilities.

A vocational evaluation provides a systematic way of determining whether a person has the necessary skills for a certain vocation (Borntbusch, 1987). Vocational evaluation, a specialty directly related to occupational therapy, focuses on discovering appropriate vocational goals. Like occupational therapists, vocational evaluators test fine motor and gross motor skills, problem solving, creative thinking, and other cognitive skills. The goals of occupational therapy programs, such as teaching independence in activities of daily living, initiating or improving work tolerance, and teaching or improving perceptual and cognitive skills, often expand vocational skills.

Vocational evaluators, rehabilitation counselors, special educators, habilitation specialists, and occupational therapists all assist in determining a client’s capacity for a certain vocation or a series of similar or different vocations. The occupational therapist must develop the skills to analyze vocationally oriented tasks and to relate this knowledge to a client’s problems and goals. Types of vocational evaluation vary depending on the comprehensiveness of the institution, the clients served, the information needs of the vocational counselors, and the backgrounds of the evaluators.

Furthermore, a wide range of vocational evaluation techniques measure job performance and evaluate factors such as motivation, initiative, ability to accept criticism, concentration, attention span, physical stamina, interpersonal relationships, and emotional maturity (Wright, 1980). Work samples consist of activities that use

Stanley Paul, MS, OTR/L, is an Occupational Therapist, Beth Abraham Hospital, 2512 Barker Avenue, New York, New York 10467.

This article was accepted for publication April 17, 1992.

The American Journal of Occupational Therapy

809
Typical intervals in test-retest reliability are several days to a week. Maturation may become a confounding variable if the interval is too long, whereas shorter intervals produce higher likelihood of practice effects (Goldstein & Tupper, 1987).

One popular and widely used work sample assessment is the Pennsylvania Bi-Manual Worksample (2nd ed.) (PBWS) (Roberts, 1969). Reliability studies of this work sample were completed with nondisabled and physically disabled populations (Roberts, 1969). The reliability compiled in 1945 by the test developers did not include persons with developmental disabilities. Additional studies are required to establish the reliability of the work sample in other clinical environments.

The research question explored by this study was whether, on readministration of the PBWS, adequate test-retest reliability could be demonstrated for adults with developmental disabilities. Adequate reliability was operationally defined as an intraclass correlation coefficient (ICC) of .80 or above. The ICC value of .80 was selected because it represents consistent performance over time (Dunn, 1989).

Method

Subjects

“Sheltered workshops are not-for-profit organizations that provide employment for disabled persons at subminimum wages under special Department of Labor Certification” (Lagomarcino, 1988, p. 312). Of 95 adults who were employees at a sheltered workshop for adults with developmental disabilities in western New York, 30 subjects were selected with a table of random numbers. The criteria for inclusion in the study were (a) mild to severe degree of mental retardation based on the Weschler Adult Intelligence Scale–Revised (Cohen, Swardlik, & Smith, 1992) or the Stanford-Binet Intelligence Scale (Cohen et al., 1992), (b) age between 20 and 60 years, and (c) lack of obvious physical disabilities or limitations that would affect voluntary movements. Seventeen subjects (57%) were men and 13 (43%) were women. The mean age for men was 35 years, for women, 34.58 years. Twenty subjects (67%) were right-handed, and 10 (33%) were left-handed. Eighteen subjects (60%) had conditions diagnosed as mental retardation (7 mild, 11 moderate), 8 (27%) had conditions diagnosed as mental retardation with secondary psychiatric diagnosis (3 mild, 5 moderate), and 4 (13%) had conditions diagnosed as Down syndrome. All subjects had adequate comprehension, verbal, and motor skills. Their IQs ranged from 24 to 97 (M = 51.40, SD = 13.95). Their ages ranged from 25 to 51 years (M = 34.73, SD = 7.13).

Instrument

By requiring the assembly and disassembly of nut and bolt units, the PBWS measures finger dexterity of both hands, gross movements of both arms, eye-hand coordination, bimanual coordination, and ability to use both hands in cooperation (Roberts, 1969). The work sample consists of a 2-ft by 8-in. wooden board with 100 holes in the center and two cut-out areas (trays), one on each side, for 105 bolts and 105 nuts. To assemble, the subject places a bolt with a nut on it in each hole. To disassemble, the subject removes the nuts from the 100 bolts and puts the nuts and bolts in the appropriate trays. During assembly, 20 holes are considered practice; 80 are scored for the test. However, no practice period is allowed for disassembly; the subject is expected to disassemble all 100 test pieces. Assembly and disassembly times are measured separately in seconds and the scores are entered on the score sheet.

Internal reliability for the whole test with the Spearman-Brown formula is .947; for the split-half method, in which the test is divided in halves and the two half-tests are correlated (Safrit, 1986), it is .897 on 112 vocational school boys (Roberts, 1969). No specific data on test-retest and interrater reliability measures are available from the PBWS test manual or other research studies that use the PBWS with persons with developmental disabilities.

Procedure

Before the test, letters were sent to the parents and legal guardians of the subjects to explain the research study and to obtain their consent. Ten subjects who were com-
petent to give their own consent had been apprised that they had full authority to participate or not to participate. I became familiar with test administration by testing five volunteer professionals and staff members affiliated with the workshop.

Initial administration of the test was conducted during a 4-day period. As directed by the test manual, each subject was seated on a 19-in. stool in front of a table 30 in. high. If the subject was extremely short or tall, the height of the table or the chair was adjusted. One hundred and five bolts and 105 nuts were placed in the proper trays (the extra 5 nuts and 5 bolts were included in case the subject dropped a few while performing the test). The instructions in the test manual were read to the subject. The subject practiced assembling 20 nuts and bolts. After this practice period, the subject assembled the remaining 80 test pieces and disassembled all 100 test pieces. A stopwatch was used to record the performance time.

The retest occurred within 7 to 10 days of the initial test administration and followed the same procedures. The 8 subjects who had conditions with a dual diagnosis were receiving psychotropic medication, but no changes in their medication occurred during the study period.

Test-retest reliability was analyzed with the intraclass correlation coefficient (ICC). A random effects one-way analysis of variance model was used to calculate the ICC value. In addition, the relationships of age, IQ, and gender to performance on the test were analyzed.

Results

The overall results revealed that the PBWS has adequate test-retest reliability for subjects with developmental disabilities (see Table 1). The ICC value of .98 corresponded with the Pearson product-moment correlation for the test-retest administration, which was also found to be .98.

A moderate positive correlation between age and performance on the test revealed that younger subjects performed faster than older ones ($r = .45, p < .05$). This finding is consistent with previous findings (Roberts, 1969). Little relationship was found, however, between IQ and performance ($r = .14, p = .05$).

No statistically significant differences in mean performance on the test were found when subjects were classified by gender or handedness. Further data analyses revealed no statistically significant difference in mean age or IQ scores for men and women or for right-handed and left-handed subjects (see Table 2).

Discussion

The results revealed a value of .98 for both the ICC and the Pearson product-moment correlation. Supplemental analysis indicated that on the average, younger workers performed better than older ones. The lack of a statistically significant difference in performance on either assembly or disassembly tasks based on gender contradicts the statement in the test manual that men have scored consistently higher on the assembly task than women have (Roberts, 1969). This difference may exist because the current study sample was small and because the earlier studies were performed primarily on nondisabled and physically disabled subjects. The men in the nondisabled population may have been more skilled at mechanical tasks than the women were because they had had experience with nuts and bolts. In the study sample, all subjects were employed in a sheltered workshop, thus men and women may have been equally familiar with tasks requiring dexterous movements of the hands and arms.

Study Limitations

The generalizability of a study is related to sample size and sample plan. Although the sample was selected randomly, all 30 subjects were from the same sheltered workshop, which limits the generalizability of the findings. A larger sample from more sheltered workshops from a wider geographical area would enhance generalizability.

### Table 1
Subjects’ Total Scores on the Pennsylvania Bi-Manual Worksample

<table>
<thead>
<tr>
<th>Trial 1/2</th>
<th>Assembly</th>
<th>Disassembly</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>4.03</td>
<td>4.00</td>
<td>8.03</td>
</tr>
<tr>
<td>Trial 2</td>
<td>3.96</td>
<td>3.95</td>
<td>7.91</td>
</tr>
</tbody>
</table>

Note: Raw scores were converted into standard scores using the $Z$ score formula: $Z = \frac{X - \bar{X}}{S}$, where $X$ = raw scores, $\bar{X}$ = mean of raw scores, and $S$ = standard deviation.

### Table 2
Results of t-Test Comparisons of Dependent Variables on the Pennsylvania Bi-Manual Worksample (PBWS) ($N = 30$)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBWS score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.38</td>
<td>.18</td>
</tr>
<tr>
<td>Handedness</td>
<td>0.823</td>
<td>.42</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.23</td>
<td>.82</td>
</tr>
<tr>
<td>Handedness</td>
<td>0.88</td>
<td>.39</td>
</tr>
<tr>
<td>IQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.16</td>
<td>.25</td>
</tr>
<tr>
<td>Handedness</td>
<td>0.60</td>
<td>.55</td>
</tr>
</tbody>
</table>

*(Roberts, 1969).*
Study Implications

This study provides an empirical basis that justifies the use of the PBWS to assess the work skills of persons with developmental disabilities. Empirical information regarding validity, that is, knowledge of how well the work sample measures specific work skills, would improve the work sample's usefulness as a vocational assessment for persons with developmental disabilities and would ensure that the work sample is providing clinically relevant and accurate information.

Future research should also examine whether the skills assessed by the PBWS are skills needed by persons with developmental disabilities in natural work environments. For example, the PBWS evaluates fine motor skills, which are necessary for many contract jobs in sheltered workshops. Recently, more adults with developmental disabilities have been getting data entry or similar jobs that require computer skills as well as fine motor skills. Therefore, a correlation study examining the performance scores in the PBWS and computer keyboard skills might be useful. Research determining the reliability of the PBWS with an interval longer than 7 days between the first and second administration would also provide useful information regarding the consistency of the derived value.

Acknowledgments

I thank my thesis committee members, K. Ottenbacher, PhD, OTR, W. Mann, PhD, OTR, and S. Nochajski, MS, OTR, for their assistance and guidance.

This study was completed as a partial requirement for the master's degree in occupational therapy from the State University of New York at Buffalo.

References


