Test-Retest Reliability of the Test of Visual Perceptual Skills With Children With Learning Disabilities

Sheri Ambacher McFall, Jean C. Deitz, Terry K. Crowe

Key Words: pediatrics • visual perception

This study examined the test-retest reliability of the Test of Visual Perceptual Skills (nonmotor) (TVPS). The sample consisted of 30 first- and second-grade children (aged 6 years through 8 years) with identified learning disabilities. The TVPS was administered on two separate occasions that were 1 to 2 weeks apart. The intraclass correlation coefficient for the total test standard scores was .81. The intraclass correlation coefficients for the subtests ranged from .33 (Sequential Memory) to .78 (Form Constancy). The primary finding from this study is that TVPS scores on the total test show adequate test-retest reliability for use in clinical settings. The scores on the subtests, however, should be used with extreme caution, as the test-retest reliability estimates were low.

Occupational therapists working with children often assess visual perceptual abilities, along with fine motor and visual motor integration skills, to gather information for use in determining service needs or measuring progress (Cook, 1991). Visual perceptual tests have been used or recommended for use with children and adolescents who have been referred because of suspected learning disabilities (Blalock, 1982; Hagerman, 1984; Johnson, 1981; Whyte, 1984). According to Gardner (1982), visual-perceptual deficits would be more common among children in a “school for learning handicapped” than in children who are “average learners” (p. 16). Therefore, assessment of visual perception in children and adolescents with learning disabilities seems warranted.

Two standardized, nonmotor tests used by occupational therapists and other professionals when assessing visual perception are the Test of Visual Perceptual Skills (TVPS) (Gardner, 1982) and the Motor-Free Visual Perceptual Test (MVPT) (Colarusso & Hammill, 1972; Crowe, 1989). The TVPS includes seven subtest areas of Visual Discrimination, Visual Memory, Visual Spatial Relationships, Visual Form Constancy, Visual Sequential Memory, Visual Figure Ground, and Visual Closure. Each subtest contains 16 items that, according to Gardner (1982), are presented in an increasing level of difficulty.

When selecting a test for use with children in clinical practice or in research, occupational therapists and other professionals need to consider the reliability of the test (Benson & Clark, 1982; Deitz, 1989). One essential type of reliability is test-retest reliability, an index of score stability over time (Anastasi, 1988) that allows therapists to be confident that score changes reflect change in the person’s performance rather than random error. Both Gardner (1982), in the TVPS test manual, and reviewers critiquing the TVPS in The Ninth Mental Measurements Yearbook (Busch-Rossnagel, 1985; Denison, 1985) indicated the need to examine test-retest reliability for this measure. This study addressed the identified need by focusing on the stability of TVPS scores. First- and second-grade children (6 to 8 years old with identified learning disabilities) were selected because these are common grades in which children with classroom difficulties are referred for occupational therapy evaluations in the public school system. The following research question was addressed: What are the test-retest reliabilities and percentages of agreement for total TVPS scores and for TVPS subtest scores for first- and second-grade children with identified learning disabilities?

Method

Sample

A convenience sample of 30 children with identified learning disabilities (25 boys and 5 girls) was tested. A child with an identified learning disability was defined as a

Sheri Ambacher McFall, MS, OTR/L, is an occupational therapist in the Everett School District, Special Services, 202 Alder, Everett, Washington 98201. At the time of this study, she was a graduate student in the Department of Rehabilitation Medicine, University of Washington, Seattle.

Jean C. Deitz, PhD, OTR/L, FaOTA, is Associate Professor, Department of Rehabilitation Medicine, University of Washington, Seattle, Washington.

Terry K. Crowe, PhD, OTR/L, is Director and Assistant Professor, Division of Occupational Therapy, Department of Orthopedics and Rehabilitation, University of New Mexico, Albuquerque, New Mexico.

This article was accepted for publication April 21, 1993.
child having a qualifying handicapping condition, which the members of the public school multidisciplinary team used to qualify the child for special education and related services. The handicapping condition of learning disability is based on the following definition as stated in the Washington Administrative Code (1988) (WAC 392-171-40):

The presence of a specific learning disability is indicated by intellectual functioning above that specified in this chapter for eligibility as mentally retarded and by a severe discrepancy between the student’s intellectual ability and academic achievement in one or more of the following areas: oral expression, listening comprehension, written expression, basic reading skill, reading comprehension, mathematics calculations, and mathematics reasoning; provided, that such performance deficit cannot be explained by visual or hearing problems, mental handicaps, mental retardation, behavioral disability, or environmental, cultural, or economic factors (p. 20).

The children were in first grade (n = 5) or second grade ($n = 25$), and ranged in age from 6 years, 9 months to 8 years, 11 months ($M = 7$ years, 7 months). All of the children in this study were white, and English was their primary language. According to parent report, only one child wore glasses and none of the children had uncorrected vision problems.

The following information was obtained from questionnaires completed by the parents. Seven of the children were reported to have had problems during or shortly after birth and three were reported to have been more than 4 weeks premature. The parents also were asked if their child had ever been identified as having one or more of the following problems (answers were recorded as yes, no, or possibly): developmental disability (yes, 7; possibly, 9); large motor problems (yes, 4; possibly, 6); neurological problems (possibly, 1); a physical disability (possibly, 4); and seizures (possibly, 1).

Academic information was obtained from each child's file at the school district's special education office (see Table 1). The classroom teacher also completed a brief questionnaire (see Table 2). The classroom teacher who completed the form was the regular education teacher unless the only teacher the child had was a special education teacher.

### Instrumentation

The Test of Visual Perceptual Skills (nonmotor) (TVPS) was used in this study. The TVPS is standardized and contains 112 items divided into 7 subtests. The individual test plates contain four to five forms that were selected without bias in regard to race, culture, gender, education, or language (Gardner, 1982). The child is given the directions, shown the test plate, and asked to indicate the correct response among four to five choices. A child may indicate his or her answer by pointing, by verbally stating the number, or by any other method established by the tester, such as using a communication board. The test is not timed. Administration needs to be done individually and requires approximately 20 to 30 min. Tables are provided in the manual for determining perceptual age, scaled scores, and percentiles for each subtest and the perceptual quotient, percentile rank, and median perceptual age for the total test (Gardner, 1982).

### Data Collection

Lists of potential subjects were obtained from two school districts in western Washington, and the parents were sent forms describing the study. With parental consent, both test and retest sessions were scheduled for each child at approximately the same time of day, either in the morning or in the afternoon. Time between testing sessions was 1 to 2 weeks ($M = 9$ days). This time allowed for schedule changes due to children's absences or class-
room requirements. Testing took place at the child's school in a small, quiet room that was the same for both testing sessions. Both the test and retest were administered by the same examiner, a pediatric occupational therapist with 5 years of clinical experience.

The TVPS was administered in accordance with the directions in the manual with all items given in each subtest until a ceiling was reached. If a child's choice was unclear to the examiner or if the child changed his or her mind, the examiner would say, "You need to give me one answer." After each testing, the child received a small token for participating.

Data Analysis

Data were analyzed in five steps. Standard scores were used for data analysis, with scaled scores used for the subtests and perceptual quotients used for the total test. The standard scores reported in the test manual (Gardner, 1982) for the total test are perceptual quotients ($M = 100, SD = 15$) and for the subtests are scaled scores that range from 1 to 19, with a scaled score of 10 considered average ($SD = 3$). These scores are commonly used clinically because they are "derived scores that transform raw scores in such a way that the set of scores always has the same mean and the same standard deviation" (Cermak, 1989, p. 107).

The first step in the data analysis process was to complete descriptive statistics for all information gathered during this study. Second, test-retest reliability was determined for each of the seven subtests and for the total test with the two-way random effects repeated measures model of the intraclass correlation coefficient. This model considers differences between test and retest scores as sources of error and provides reliability coefficients that can be generalized to another examiner (Tinsley & Weiss, 1975). A computer program written by Paulson and Trevisan (1990) was used to run the intraclass correlation. With this program, formula 2 was used. Third, standard errors of measurement were calculated with the square root of the residual mean squares (Fleiss, 1986; M. Trevisan, personal communication, March 25, 1993). Fourth, Pearson product-moment correlations were determined for the total test and subtests to facilitate comparison of reliability estimates between the TVPS and the MVPT. In addition, paired t-tests were computed to determine any significant differences between mean scores on test and retest for the total test and the subtests.

Analyzing the percent of agreement between the test and retest for each of the seven subtests and for the total test was the fifth step in data analysis. Analysis of the difference between the standard scores obtained during the two testing sessions shows that the clinical implications of the differences are evident. The magnitudes of difference (larger score minus smaller score) between individual test and retest standard scores for both the total test score and the subtest scores were calculated to determine the actual differences between the scores.

Results

Descriptive statistics for the total test and subtest standard scores are reported in Table 3. Most scores show a tendency toward improvement on the retest. The mean improvement was 5.9 points for the total test standard scores and ranged from 0.0 to 19.0 points for the subtests. According to paired t-tests examining the differences between the group means on test and retest, significant differences ($p < .05$) occurred for the total test scores and for the subtests of Visual Closure and Visual Memory (see Table 4).

Intraclass correlation coefficients (ICCs), standard errors of measurement, and Pearson product-moment

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total test</td>
<td>97.2</td>
<td>92.0</td>
<td>16.5</td>
<td>64.0</td>
<td>129.0</td>
</tr>
<tr>
<td>Test</td>
<td>103.1</td>
<td>102.0</td>
<td>19.2</td>
<td>67.0</td>
<td>147.0</td>
</tr>
<tr>
<td>Subtests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual discrimination</td>
<td>11.9</td>
<td>12.5</td>
<td>3.1</td>
<td>5.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Test</td>
<td>13.0</td>
<td>12.0</td>
<td>3.7</td>
<td>5.9</td>
<td>19.0</td>
</tr>
<tr>
<td>Visual memory</td>
<td>9.4</td>
<td>10.0</td>
<td>3.7</td>
<td>5.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Test</td>
<td>10.5</td>
<td>10.0</td>
<td>3.1</td>
<td>3.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Spatial relations</td>
<td>10.4</td>
<td>10.5</td>
<td>4.1</td>
<td>2.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Test</td>
<td>11.1</td>
<td>10.0</td>
<td>4.3</td>
<td>4.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Form constancy</td>
<td>8.7</td>
<td>9.0</td>
<td>3.1</td>
<td>3.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Test</td>
<td>9.4</td>
<td>9.5</td>
<td>3.3</td>
<td>1.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Sequential memory</td>
<td>8.4</td>
<td>8.0</td>
<td>2.6</td>
<td>4.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Test</td>
<td>8.4</td>
<td>8.0</td>
<td>2.6</td>
<td>4.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Figure ground</td>
<td>9.2</td>
<td>8.5</td>
<td>3.8</td>
<td>2.0</td>
<td>17.0</td>
</tr>
<tr>
<td>Test</td>
<td>9.3</td>
<td>9.0</td>
<td>3.7</td>
<td>4.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Visual closure</td>
<td>9.5</td>
<td>9.5</td>
<td>3.3</td>
<td>4.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Test</td>
<td>11.4</td>
<td>11.0</td>
<td>3.8</td>
<td>4.0</td>
<td>19.0</td>
</tr>
</tbody>
</table>

| Total test | -3.28 | .005* |
| Subtests | | | |
| Visual discrimination | -1.89 | .060 |
| Visual memory | -2.41 | .012* |
| Spatial relations | -1.12 | .271 |
| Form constancy | -1.76 | .089 |
| Sequential memory | -0.01 | 1.000 |
| Figure ground | -1.11 | 911 |
| Visual closure | -3.17 | .003* |

*p < .05
correlation coefficients for test-retest reliability for total test standard scores and subtest standard scores are presented in Table 5. Magnitudes of difference between total test standard scores for test and retest are reported in Table 6. Approximately 77% of the subjects obtained a higher score on retest total scores as compared to test total scores. Magnitudes of difference for the subtest standard scores are reported in Table 7. For the magnitudes of difference, the percentages of all subjects who improved, showed no change, or declined in standard scores for each subtest are shown in Table 8. For six of the seven subtests, 50% or more of the subjects showed improvement on the retest.

Discussion

The test-retest ICC obtained in this study for the TVPS total test was .81. For the subtests, the ICCs ranged from .55 (Visual discrimination) to .78 (Form Constancy). These results suggest that the stability of scores for children with an identified learning disability is adequate for the total test but is generally poor for the subtests. Therefore, less confidence should be placed in the stability of subtest scores.

When reliability was examined for the TVPS with Pearson product-moment correlation coefficients, higher coefficients were reported than with the ICC. However, paired t-tests indicated a significant difference between test and retest scores for the total test scores and subtest scores of Visual Closure and Visual Memory. This difference appears to reflect a practice effect. Intraclass correlation coefficients consider this type of systematic change to be part of measurement error; therefore, intraclass correlation coefficients probably provide better estimates of test-retest reliability on this measure.

In addition to the TVPS, occupational therapists also use the MVPT to assess visual perception. For the MVPT, Pearson product-moment test-retest coefficients ranged from a low of .77 (4-year-olds) to a high of .83 (6-year-olds), with a coefficient of .81 for the total sample of 162 subjects who were considered to be typically developing.

Professionals who have both tests available may...
need to weigh several factors when deciding which test to use. First, it may take less time to use the MVPT, which has only 36 items. Second, obtaining subtest scores has been considered a benefit of the TVPS (Gardner, 1982). However, because of low reliability estimates, extreme caution is indicated when using these scores. Third, the TVPS provides normative data for a broader age range (4 through 12 years) as compared to the MVPT (4 through 8 years).

Score stability on the Figure Ground subtest of the TVPS also can be compared to the Figure Ground subtest of the Southern California Sensory Integration and Praxis Test (Ayres, 1989). A test-retest reliability study was completed for the latter subtest using a sample of 41 children with learning disabilities and a mean age of 6.5 years (Ayres, 1989). The reliability estimate for this measure \( r = .54 \) was slightly lower than that for the Figure Ground subtest of the TVPS \( r = .63 \).

Percentage agreement also was used to examine the stability of test scores. Examining percentage agreement allows a closer look at the comparison between two sets of scores based on absolute values. Clinically, even small point differences in scores may be enough to influence service need decisions or potential programming decisions. Examination of the percentage changes based on the magnitudes of difference shows that 50% or more of the subjects improved on retest for all of the subtests except Sequential Memory (see Table 8). On the basis of these results, some instability appears to be evident in the scores on the subtests.

When the magnitudes of difference between the total test scores for test and retest were examined, 76.7% of the subjects obtained a higher score on the retest. Examination of the total test scores (see Table 6), shows that 5 of the 30 subjects obtained retest scores 15 points or more higher than their test scores and 1 subject obtained a score 16 points lower on the retest. Therefore, 20% of the subjects’ scores changed by one standard deviation or more between test and retest. Such a change could affect decisions that may be made about a child’s visual perceptual abilities.

The ability to generalize the results of this study is limited by two factors. First, only one rater was used for both testing sessions. The increased familiarity with this rater at the second testing session may have caused some children to perform differently at the second testing even with the attempt to keep testing consistent. Results are therefore only generalizable to situations in which the same tester completes the initial testing and provides treatment before retesting the child at a later time. This is a common situation in the public school system. Second, the sample was small and only represented two school districts in one geographical area. If the current study were replicated with a larger sample from a more diverse area, therapists could have even more confidence in the stability of the reliability coefficients generated.

According to some authors (Blalock, 1982; Hagerman, 1984; Johnson, 1981; Whyte, 1984), children with learning disabilities may have suspected deficits in visual perception. When completing the standardization and norming process for the TVPS, Gardner (1982) compared 45 students who attended a school for the learning handicapped with a matched group from the standardization sample. The children were matched on the basis of age, gender, and race. The mean age of the children in his study was 10.11 years. On the basis of ANOVA, Gardner (1982) noted lower performances for the group of children with learning disabilities.

The children in the current study qualified as learning disabled and had a mean age of 7 years, 7 months. Their total scores reflect overall performances near the mean (test: \( M = 97.2 \), retest: \( M = 103.1 \)). Therefore, the 6- to 8-year-old children with identified learning disabilities who participated in this study showed average performances on this visual perceptual test. The discrepancy between the performance on the TVPS of subjects in our sample as compared to the performance of subjects in Gardner’s sample leads to the question of whether children with learning disabilities have visual perceptual deficits as measured by the TVPS. Further research is warranted in this area, possibly examining the visual perceptual skills of subgroups of children of specific ages whose learning disabilities are reflected in specific types of academic performance deficits.

Conclusion

Results of this study suggest that when the TVPS is administered to 6- to 8-year-old children with identified learning disabilities, the stability of total test scores can be considered adequate. When used for determining progress, these scores should be interpreted cautiously, as they showed an overall increase on the retest even with no intervention. Test-retest reliability estimates for the subtests ranged from low to borderline acceptable; therefore, these scores are of limited use for determining service needs and documenting progress.

Acknowledgments

We thank the children from the Everett and Northshore School Districts who participated in the study. This study was supported in part by the Paula M. Carman Memorial Fellowship Fund.

References


