
Reliability of Two Visual–Perceptual Tests for Children With Cerebral Palsy

Li-ting Tsai, Keh-chung Lin, Hua-fang Liao, Ching-lin Hsieh

KEY WORDS

- cerebral palsy
- child
- psychometrics
- reproducibility of results
- visual perception

OBJECTIVE. We examined the reliability of the Motor-Free Visual Perception Test–Revised (MVPT–R) and the Test of Visual–Perceptual Skills–Revised (TVPS–R) in 52 children with cerebral palsy.

METHOD. The MVPT–R and TVPS–R were initially assessed by a single rater and then randomly assigned to the same rater or another rater after 6 to 14 days.

RESULTS. For scale-level reliability of both tests, the intraclass correlation coefficients (ICCs) were high. The smallest real differences (SRDs) were largely acceptable. The ICCs of most subscale-level reliability of the TVPS–R were satisfactory. The SRD of each TVPS–R subscale score was notable (>29%). The internal consistency was high for both tests.

CONCLUSION. The scale-level reliabilities of both tests were satisfactory in children with cerebral palsy. Although most subscales of the TVPS–R had acceptable test–retest or interrater agreement, all subscales of the TVPS–R had large SRDs, limiting their clinical utility.

Tsai, L.-T., Lin, K.-C., Liao, H.-F., & Hsieh, C.-L. (2009). Reliability of two visual–perceptual tests for children with cerebral palsy. *American Journal of Occupational Therapy, 63*, 473–480.

Li-ting Tsai, MS, is PhD Student, Taipei Parents' Association for the Visually Impaired, National Taiwan University Hospital, Taipei.

Keh-chung Lin, ScD, is Associate Professor and Chair, School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, and Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei.

Hua-fang Liao, MS, is Associate Professor, School and Graduate Institute of Physical Therapy, College of Medicine, National Taiwan University, Taipei, and Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei.

Ching-lin Hsieh, PhD, is Professor, School of Occupational Therapy, College of Medicine, National Taiwan University, 4F, No 17, Shuijous Road, Taipei 100, Taiwan; clhsieh@ntu.edu.tw

Cerebral palsy, the most common cause of physical disability in children, causes disorders of movement and posture because of a defect or lesion of the immature brain (Bax et al., 2005). Although the clinical picture of cerebral palsy is focused on motor dysfunction, brain damage from cerebral palsy is also responsible for other defects. Among these defects, visual–perceptual problems are common in children with cerebral palsy (Ito et al., 1996; Koeda & Takeshita, 1992) and can affect children's reading ability (Kozeis et al., 2006) and learning (Pirila et al., 2004). To make accurate clinical judgments, plan treatment, and measure the outcome of intervention, a test of visual perception with sufficient reliability in children with cerebral palsy is warranted.

Reliability, the degree to which a measurement is consistent and free from measurement error, is the essential and first consideration in the examination of the psychometric properties of a test (Portney & Watkins, 2000). Relative and absolute statistical indexes have been suggested to fully establish a test's reliability (Flansbjerg, Holmback, Downham, Patten, & Lexell, 2005; Rankin & Stokes, 1998). Relative indexes, such as the *intraclass correlation coefficient (ICC)*, are used to estimate the level of agreement between repeated measurements (for group comparison or individual measurements over time; Flansbjerg et al., 2005; Portney & Watkins, 2000) and to compare the degree of reliability between tests. Absolute indexes are used to quantify the extent of measurement error of a test (Flansbjerg et al., 2005). *Smallest real difference (SRD)*, also called *minimal detectable change*, is one of the absolute indexes and has been developed to indicate a real change (beyond measurement error) for a person (Flansbjerg et al., 2005; Haley & Fragala-Pinkham, 2006). Therefore,

SRD, not ICC, can provide a specific threshold for clinicians to determine whether each of their clients has improved or deteriorated. A useful test should have a low SRD value to ensure that a minimum change is beyond measurement error.

Two standardized, nonmotor visual–perceptual tests commonly used by pediatric therapists (Burtner et al., 1997) are the Motor-Free Visual Perception Test–Revised (MVPT–R; Colarusso & Hammill, 1996) and the Test of Visual-Perceptual Skills–Revised (TVPS–R; Gardner, 1996). Both nonmotor tests consist of predrawn configurations and designs. Children are required to indicate the correct answer out of four or five choices by any means (e.g., by pointing or verbal indications). The TVPS–R consists of seven subscales. The total scores and subscale scores of the TVPS–R are usually used in research and clinical settings. However, although the MVPT–R contains five subareas of visual–perceptual skills, only the use of total scores is suggested by the authors (Colarusso & Hammill, 1996). Because nonmotor visual–perceptual tests could minimize the motor effect on visual–perceptual performance, these tests are suitable for children with motor impairment (e.g., cerebral palsy). The reliabilities of the MVPT–R and TVPS–R have been validated in normative samples and children with learning disabilities (Burtner, Qualls, Ortega, Morris, & Scott, 2002; Chan & Chow, 2005; Colarusso & Hammill, 1996; McFall, Deitz, & Crowe, 1993). However, to our knowledge, no articles that evaluate the psychometric properties of the tests in children with cerebral palsy have been published. Moreover, because the psychometric properties of tests are generally sample dependent, the properties in one population cannot automatically be generalized to other populations (Gliner, Morgan, & Harmon, 2001; Portney & Watkins, 2000). It is thus necessary to determine the reliabilities of the MVPT–R and TVPS–R before their application to children with cerebral palsy.

The purposes of this study were to examine the test–retest reliabilities, interrater reliabilities, and internal consistency of the MVPT–R and the TVPS–R in children with cerebral palsy. ICC and SRD were used to analyze the scale-level reliabilities of both tests and subscale-level reliabilities of the TVPS–R. In addition, because item-level reliability can provide information about whether an item needs to be revised or replaced, the item-level reliabilities of both tests were also examined in this study.

Method

Participants

A convenience sample of children with cerebral palsy was recruited from regular kindergarten and elementary educa-

tion classrooms by special educationists or from rehabilitation units by occupational therapists in Taipei, Taiwan, from November 1, 2005, to December 31, 2005. The inclusion criteria were as follows: (1) a diagnosis of cerebral palsy, (2) ages 5 to 8, and (3) an ability to follow general oral instructions. We excluded children as follows: (1) if they could not follow the instructions of the MVPT–R and TVPS–R and (2) if they had poor visual acuity as indicated by the Teller Acuity Cards (TAC; Teller, 1990) assessments at 38 cm of less than 6.50 cycle/cm.

Sixty children with cerebral palsy (34 boys, 26 girls) met the inclusion criteria. Five children (2 boys, 3 girls) were excluded because they could not follow the necessary instructions of the MVPT–R or the TVPS–R. Therefore, 55 children were eligible for recruitment. Three children were not recruited because they could not finish the whole assessment, and the remaining 52 children (31 boys, 21 girls) participated in this study.

All the procedures for this study were approved by a local institutional review board. Parents gave written informed consent for their child's participation.

Procedures

Three raters (all occupational therapists) participated in this study. The first rater (Li-ting Tsai) collected demographic and clinical information, including binocular grating acuity (Teller, 1990), neuromotor abnormalities, functional motor abnormalities, and associated impairment (Bax et al., 2005). The other two raters administered the reliability investigation. Before the initiation of this study, rater training was administered to reduce the raters' bias.

The reliability investigation consisted of two sessions for the test–retest and interrater reliability investigation. In Session 1, the MVPT–R and TVPS–R were administered by an occupational therapist. Approximately 7 days later, the participants were randomly and equally assigned to the same occupational therapist (for the test–retest reliability investigation) or the other occupational therapist (for the interrater reliability investigation) in Session 2. The timeframe for this study was chosen for two reasons. First, we hoped to avoid a learning effect or memory effect after a good deal of practicing similar tasks. The second reason was a practical issue. Most children received treatment once a week. Our design would increase parents' motivation to allow participation.

Testing sequences of the MVPT–R and TVPS–R were counterbalanced. That is, if a child (Participant 1) started with the MVPT–R during the first session, then the next recruited child (Participant 2) would start with the TVPS–R. In both sessions for each child, the testing sequences were the same. The testing conditions, including environment and time, were as constant as possible on the two separate

occasions. Participants were allowed to rest during testing when necessary.

Instruments

The MVPT–R (the second version) is designed to assess the visual–perceptual strengths and weaknesses of children ages 4 years through 11 years 11 months. The MVPT–R consists of 40 items, divided into five types: spatial relationships, visual memory, visual discrimination, figure ground, and visual closure (Colarusso & Hammill, 1996). The authors suggested using the total scores of the MVPT–R, which ranged from 0 to 40 points. The test–retest reliability coefficient ranged from .70 to .83 in a normative sample and children with learning disability (Burtner et al., 2002; Colarusso & Hammill, 1996). Internal consistency ranged from .81 to .84 at various age levels in a normative sample (Colarusso & Hammill, 1996). Although the third version (the MVPT–3) has been developed recently, the first 40 items of the MVPT–3 are exactly the same as the 40 items of the MVPT–R. Thus, using either the MVPT–R or the MVPT–3 is the same for our participants.

The TVPS–R is also designed to assess the visual–perceptual strengths and weaknesses of children ages 4 years through 12 years 11 months (Gardner, 1996). The TVPS–R consists of 112 items grouped in seven subscales (i.e., visual discrimination, visual memory, visual–spatial relationships, visual form constancy, visual sequential memory, visual figure ground, and visual closure). The authors suggested using both the total scale, with scores ranging from 0 to 112 points, and the subscales of the TVPS–R, with scores ranging from 0 to 16 points. The reliability is adequate for the total scale but not for the subscales (ICCs = .33 to .78) or average individual items in children with learning disability and Chinese preschoolers (Chan & Chow, 2005; McFall et al., 1993). In addition, the internal consistency of the total scale was satisfactory, but it was not acceptable for most of the subscales of the TVPS–R across all age levels in the standardization sample (Gardner, 1996).

Data and Statistical Analysis

Test–Retest and Interrater Reliability. The scale level of the MVPT–R and TVPS–R and subscale level of the TVPS–R test–retest and interrater reliabilities were analyzed with ICC and SRD. The item-level test–retest and interrater reliabilities of the MVPT–R and TVPS–R were examined with the percentage agreement.

In the scale-level and subscale-level analysis, the raw scores of the total scale of the MVPT–R and TVPS–R and the subscale of the TVPS–R were used in the analysis. Agreement between measurements was analyzed by the random effect of ICC Model 2. An ICC value >.90 indicated

excellent agreement and appropriateness for individual comparisons. An ICC value >.70 indicated acceptable agreement and appropriateness for group comparisons (Aaronson et al., 2002). The SRD was used to quantify measurement errors because of chance variation in measurements (Haley & Fragala-Pinkham, 2006). The SRD was calculated using the following formula (Beckerman et al., 2001; Flansbjerg et al., 2005): $SRD = 1.96 \times \sqrt{2} \times SEM$.

The standard error of measurement (SEM) was calculated as the square root of the within-participant variance. For each person, the difference between consecutive measurements should be at least $1.96 \times SEM \times \sqrt{2}$ to indicate 95% confidence of a real difference between two scores (1.96 because of the 95% confidence, $\sqrt{2}$ because of the difference of two variances; Beckerman et al., 2001). In item-level analysis, a percentage of agreement >.80 indicated high agreement.

Internal Consistency. The internal consistencies of the MVPT–R and TVPS–R were determined by Cronbach's alpha. The data of the first assessment were used for the internal consistency investigation. An alpha coefficient >.70 was considered adequate for group comparison (Aaronson et al., 2002).

Results

Descriptive Statistics

Fifty-two children with cerebral palsy finished the whole assessment. The chronological ages of the 52 participants (31 boys, 21 girls) ranged from 5 years 5 months to 8 years 9 months (mean age = 6 years 11 months; standard deviation [*SD*] = 1 year). The neuromotor abnormalities were spastic type ($n = 48$), dyskinetic type ($n = 2$), and ataxic type ($n = 2$). Functional motor abnormalities were assessed using the Gross Motor Function Classification System (Palisano et al., 1997) and the Manual Ability Classification System (Eliasson et al., 2006; see Table 1). Three-quarters of the participants had visual impairment, such as myopia, hyperopia, or strabismus. Table 1 shows characteristics of the participants. Table 2 shows the raw scores of the MVPT–R and TVPS–R in Session 1.

Both the test–retest and interrater reliability investigation groups had 26 participants. There were no significant differences in the characteristics between intrarater and interrater groups, including age, gender, visual acuity, education, and functional motor ability ($p > .05$). The intervals between Sessions 1 and 2 ranged from 6 to 14 days (mean [*M*] = 7.6 days; *SD* = 1.8 days). A total of 48 children were retested within 6 to 9 days. Only 4 children were retested after 12 or more days.

Table 1. Characteristics of the Participants (N = 52)

Participants	<i>n</i>	%
Male	31	60
Female	21	40
Type of cerebral palsy		
Spastic	48	92
Bilateral	37	71
Hemiplegia	11	21
Dyskinetic	2	4
Ataxic	2	4
Associated impairments		
Seizures	7	13
Hearing impairment	3	6
Visual impairment	39	75
GMFCS level		
Level I	10	19
Level II	21	40
Level III	11	21
Level IV	5	10
Level V	5	10
MACS level		
Level I	15	29
Level II	19	37
Level III	13	25
Level IV	3	6
Level V	2	4

Note. GMFCS = Gross Motor Function Classification System; MACS = Manual Ability Classification System.

Test-Retest Reliability

Scale Level and Subscale Level. The results of the scale-level and subscale-level ICC analyses are shown in Table 3. The test-retest agreement for the total scores of the MVPT-R was excellent, with an ICC of .96 (95% confidence interval [CI] = .90-.98). The agreement for the total scores of the TVPS-R was also excellent, with an ICC of .97 (95% CI = .93-.99). The ICC values for the subscale scores of the TVPS-R were high, with ICCs ranging from .76 to .92.

Table 3 shows the results of SRD analysis. The SRD for the scale level was 3.9 (9.6% of the total score) for the MVPT-R and 13.5 (12% of the total score) for the TVPS-R. The SRD for the subscale level of the TVPS-R ranged from 3.5 (29.2% of the total score on the sequential memory subscale) to 5.5 (45.8% of the total score on the visual form consistency subscale).

Item Level. The item-level test-retest reliability of the MVPT-R is presented in Table 4, and the item-level test-retest reliability of the TVPS-R is shown in Table 5. In the MVPT-R, the percentage agreement ranged from 42% to 96% ($M = 71.3\%$), and 11 items (28%) showed good agreement ($\geq 80\%$). In the TVPS-R, the percentage agreement

Table 2. Descriptive Data of First Raw Scores of the Motor-Free Visual Perception Test-Revised (MVPT-R) and the Test of Visual-Perceptual Skills-Revised (TVPS-R)

Test	Mean (<i>SD</i>)	Range
MVPT-R total scale	25.4 (7.2)	8-37
TVPS-R total scale	50.9 (26.5)	9-100
TVPS-R subscale		
Visual discrimination	9.1 (4.8)	1-16
Visual memory	8.7 (4.1)	1-16
Visual-spatial relationships	9.2 (5.6)	0-16
Visual form constancy	6.0 (4.2)	0-16
Visual sequential memory	5.7 (4.8)	0-15
Visual figure ground	6.2 (4.0)	1-14
Visual closure	6.0 (4.1)	0-16

Note. $N = 52$. SD = standard deviation.

ranged from 46% to 100% ($M = 74.7\%$), and 37 items (33%) showed good agreement.

Interrater Reliability

Scale Level and Subscale Level. Table 3 shows the scale-level reliabilities of the MVPT-R and TVPS-R and the subscale-level reliabilities of the TVPS-R. The interrater agreements for the total scores of the MVPT-R and the total scores of the TVPS-R were excellent, with an ICC of .92 (95% CI = .82-.96) for the MVPT-R and .93 (95% CI = .85-.97) for the TVPS-R. The ICC for the subscale scores of the TVPS-R was high, with ICCs ranging from .74 to .89, except for the visual sequential memory subscale (ICC = .63).

The SRD was 5.3 (13.3%) for the total scores of the MVPT-R and 17.3 (14.4%) for the total scores of the TVPS-R (see Table 3). The SRD for the subscale scores of

Table 3. Test-Retest ($n = 26$) and Interrater ($n = 26$) Reliability of the Motor-Free Visual Perception Test-Revised (MVPT-R) and the Test of Visual-Perceptual Skills-Revised (TVPS-R) at Scale Level and Subscale Level

Test	Test-Retest ($n = 26$)		Interrater ($n = 26$)	
	ICC (95% CI)	SRD	ICC (95% CI)	SRD
MVPT-R total scale	.96 (.90-.98)	3.9	.92 (.82-.96)	5.3
TVPS-R total scale	.97 (.93-.99)	13.5	.93 (.85-.97)	17.3
TVPS-R subscale				
Visual discrimination	.90 (.78-.96)	4.2	.74 (.43-.88)	5.2
Visual memory	.87 (.71-.94)	4.1	.76 (.46-.89)	4.3
Visual-spatial relationships	.90 (.79-.96)	4.8	.81 (.59-.92)	5.5
Visual form constancy	.76 (.47-.89)	5.5	.89 (.76-.95)	3.7
Visual sequential memory	.92 (.83-.96)	3.5	.63 (.18-.84)	7.1
Visual figure ground	.76 (.46-.89)	5.5	.87 (.71-.94)	4.0

Note. ICC = intraclass correlation coefficient; CI = confidence interval; SRD = smallest real difference.

Table 4. Test-Retest ($n = 26$) and Interrater ($n = 26$) Reliability of the Motor-Free Visual Perception Test-Revised at Item Level (Percentage of Agreement)

Item	Visual Discrimination		Form Constancy		Visual Memory		Visual Closure		Spatial Orientation	
	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater
1	96	92	77	80	73	85	50	54	92	92
2	81	92	81	85	81	92	77	62	73	88
3	85	88	65	69	77	77	69	69	58	65
4	96	85	73	65	81	69	62	69	73	88
5	77	69	81	69	62	69	62	62	69	77
6	81	77	—	—	73	88	73	65	62	58
7	85	62	—	—	54	50	69	50	—	—
8	69	73	—	—	50	69	69	88	—	—
9	—	—	—	—	—	—	69	69	—	—
10	—	—	—	—	—	—	50	65	—	—
11	—	—	—	—	—	—	42	88	—	—
12	—	—	—	—	—	—	58	69	—	—
13	—	—	—	—	—	—	77	58	—	—

Note. — = not applicable.

the TVPS-R ranged from 3.7 (30.8%) for the visual form consistency subscale to 7.1 (59.2%) for the visual sequential memory subscale.

Item Level. Table 4 shows that the percentage agreement of each item of the MVPT-R ranged from 50% to 92% ($M = 73.7\%$), with 14 items (35%) indicating good agreement. Table 5 shows that the percentage agreement of each item of the TVPS-R ranged from 50% to 100% ($M = 75.4\%$), with 36 items (32%) indicating good agreement.

Internal Consistency

The internal consistency of the MVPT-R and TVPS-R was good, with Cronbach's alpha of .87 for the MVPT-R and

.98 for the TVPS-R. At the subscale level of the TVPS-R, the Cronbach's alpha was also high, ranging from .87 to .94.

Discussion

Both absolute and relative indexes are suggested to examine the reliability of a test to provide sufficient information (Flansbjerg et al., 2005; Rankin & Stokes, 1998). To the best of our knowledge, this study is the first to use absolute and relative statistical methods to assess the reliabilities of the MVPT-R and TVPS-R in children with cerebral palsy. In addition to the scale-level reliability analysis of the MVPT-R and TVPS-R, the subscale-level reliability of the TVPS-R

Table 5. Test-Retest ($n = 26$) and Interrater ($n = 26$) Reliability of the Test of Visual-Perceptual Skills-Revised at Item Level (Percentage of Agreement)

Item	Visual Discrimination		Visual Memory		Visual-Spatial Relationship		Visual Form Consistency		Visual Sequential Memory		Visual Figure Ground		Visual Closure	
	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater	Test-Retest	Interrater
1	85	96	100	100	85	88	81	88	85	81	85	85	81	92
2	85	88	77	62	58	69	73	85	58	62	69	92	65	77
3	88	85	69	88	85	88	58	73	73	50	62	77	65	85
4	73	77	81	81	58	73	65	69	73	62	77	81	69	69
5	77	81	69	62	73	62	54	81	69	77	77	77	77	77
6	73	62	46	77	77	77	73	85	85	65	65	81	62	65
7	65	69	69	88	69	62	73	69	73	73	73	73	65	77
8	77	69	77	62	65	88	62	62	77	73	85	73	62	62
9	65	77	69	77	69	85	81	88	81	81	69	69	77	62
10	73	65	58	62	58	69	69	77	88	73	73	77	73	81
11	77	58	81	62	81	73	88	77	81	73	69	77	81	73
12	81	65	81	65	62	73	77	73	92	73	77	85	96	92
13	85	54	65	85	92	88	65	65	81	81	73	88	85	88
14	69	69	65	58	73	77	77	73	77	73	69	85	88	88
15	65	65	77	77	81	62	85	73	81	73	88	92	92	88
16	85	69	65	73	85	58	77	81	73	69	85	73	85	96

and item-level reliability of the MVPT-R and TVPS-R were simultaneously examined. These results can provide researchers and clinicians with comprehensive and useful information for further selection and application of both tests in children with cerebral palsy.

Our results indicate that the total score test-retest and interrater reliabilities of the MVPT-R and TVPS-R were generally acceptable. The total score test-retest agreements and interrater agreements of both tests were excellent, with ICC $>.90$. In terms of measurement error, the SRDs of both tests were $\leq 15\%$ of their corresponding total scores, indicating that a low threshold is needed to detect the effects of treatment on a person. These observations suggest that the total scales of the two tests can be used to monitor change in visual perception of a group or an individual. It is noted that the MVPT-R has only 40 items, whereas the TVPS-R has 112 items. Thus, the MVPT-R appears to be more efficient than the TVPS-R.

Although the subscale scores of the TVPS-R have usually been used to assess visual-perceptual skills in children with cerebral palsy (Kulp, Edwards, & Mitchell, 2002; Menken, Cermak, & Fisher, 1987), our results indicate that researchers and clinicians should interpret the subscale scores of the TVPS-R for children with cerebral palsy with caution. Only three of the seven subscales of the TVPS-R had satisfactory test-retest agreement (ICC $>.90$) for individual comparison. The test-retest and interrater reliability of all the subscales of the TVPS-R were good (ICC $>.70$) for group comparison, except for the interrater agreement of the sequential memory subscale. However, the SRDs of all subscales were unsatisfactory ($>29\%$ of the total score of each subscale). These results should serve as a warning to potential users that they must tolerate the high thresholds of each subscale of the TVPS-R to detect real change of individual visual-perceptual skills in children with cerebral palsy.

For the clinical utility of SRD, the change scores in the visual-perceptual tests should be higher than the SRD to determine whether a child benefits from visual-perceptual training. For example, if a child with cerebral palsy scores 20 points on the total scale of the MVPT-R (SRD = 3.9 for test-retest and 5.3 for interrater) at the first assessment, this child should score at least 24 points (both assessments administered by the same rater) or 26 points (assessments administered by two different raters) on the next assessment. Then the clinician can claim the change is beyond measurement error with a confidence level greater than 95%. Conversely, if the change scores are lower than the SRD, this change would likely be caused by measurement error.

Item-level reliability is useful for test developers to determine whether an item needs to be replaced or revised.

Modest item reliability indicates that the content or the administration of an item should be improved. The percentage of agreement was poor ($<.80$) in two-thirds of MVPT-R and TVPS-R items. The reason for such a high disagreement may be the learning effect, particularly because of item similarities in the tests, both of which were tested on the children at the same time. In addition, the lengthy administration may affect the participant's performance because attention problems are commonly noticed in children with cerebral palsy (McDonough & Cohen, 1982). However, behavioral observations of our participants suggest that revision and increased practice in some items of the MVPT-R and TVPS-R are warranted. For example, some of our participants appeared not to understand how to respond to the items of the visual closure subscale of the MVPT-R and TVPS-R, even after finishing examples of the visual closure subscale. Therefore, these participants often showed improvement when retested on the items of the visual closure subscale (data not shown). Thus, it is suggested that increasing the numbers of items for practice would enable children to become familiar with the test and thus might improve the item-level agreement.

The Cronbach's alpha of the total scales of the MVPT-R and TVPS-R and the subscales of the TVPS-R were high, indicating that the homogeneity for the total scales of the MVPT-R and TVPS-R and the subscales of the TVPS-R are satisfactory. However, when the items of a test show overly high levels of correlation (e.g., for the total scale of the TVPS-R, Cronbach's alpha = .97), some of these items may be redundant. Redundant items reduce the efficiency of a test and place unnecessary burden on both clients and clinicians. These observations indicate that further studies to shorten the TVPS-R are needed.

Limitations

Three limitations might concern readers. First, the sample size was not large, with one result being that only a few children with dyskinetic cerebral palsy or ataxic cerebral palsy were included. Because the types of cerebral palsy might affect children's performance on the MVPT-R and TVPS-R (Marozas & May, 1986), further studies with larger sample sizes representing all types of cerebral palsy are necessary to increase the external validity of the study results. Second, we did not examine the responsiveness of the MVPT-R and TVPS-R. The extents of the abilities of both tests to detect change in the visual perception of children with cerebral palsy remains to be investigated. Third, we adopted a 6-to-14-day period for test-retest and interrater reliability examination to avoid the learning and memory effects of the children after several practices on similar tasks.

Some authors have adopted longer periods for their reliability studies. For example, the average period of test–retest reliability was 20 days for the MVPT–R and 34 days for the MVPT–3 (Colarusso & Hammill, 1996, 2003). Although we found sufficient test–retest and interrater reliability of the TVPS–R and MVPT–R, how the timeframe affects the results of test–retest and interrater reliability remains to be investigated.

Conclusion

In summary, the results of this study indicate that the scale-level reliabilities of the MVPT–R and TVPS–R are acceptable in groups or in individuals. All subscales of the TVPS–R had acceptable internal consistency but large SRDs. In addition, only one-third of the items of the MVPT–R and TVPS–R had good item-level agreement. These results indicate that the total scales of the MVPT–R and TVPS–R can be used at the individual or group level. Moreover, the subscales of the TVPS–R can be used in group comparisons for children with cerebral palsy, but caution is necessary when interpreting the subscale scores of the TVPS–R in individual comparisons because of the large measurement error. Revision of the content or administration of the items with modest reliability, reduction of the number of items, and further validation of both tests in children with cerebral palsy all need to be accomplished in the future. ▲

Acknowledgments

We thank pediatric ophthalmologist Ai-Hou Wang from the Pediatric Ophthalmology and Adult Strabismus Department of Ophthalmology, National Taiwan University Hospital, for suggesting the exclusion criteria of the Teller Acuity Cards.

References

- Aaronson, N., Alonso, J., Burnam, A., Lohr, K. N., Patrick, D. L., & Perrin, E. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193–205.
- Bax, M., Goldstein, M., Rosenbaum, P., Leviton, A., Paneth, N., Dan, B., et al. (2005). Proposed definition and classification of cerebral palsy, April 2005. *Developmental Medicine and Child Neurology, 47*, 571–576.
- Beckerman, H., Roebroek, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research, 10*, 571–578.
- Burtner, P. A., Qualls, C., Ortega, S. G., Morris, C. G., & Scott, K. (2002). Test–retest reliability of the Motor-Free Visual Perception Test–Revised (MVPT–R) in children with and without learning disabilities. *Physical and Occupational Therapy in Pediatrics, 22*, 23–36.
- Burtner, P. A., Wilhite, C., Bordegaray, J., Moedl, D., Roe, R. J., & Savage, A. R. (1997). Critical review of visual–perceptual tests frequently administered by pediatric therapists. *Physical and Occupational Therapy in Pediatrics, 17*, 39–61.
- Chan, P. L. C., & Chow, S. M. K. (2005). Reliability and validity of the Test of Visual–Perceptual Skills (Non-Motor)–Revised for Chinese preschoolers. *American Journal of Occupational Therapy, 59*, 369–376.
- Colarusso, R. P., & Hammill, D. D. (1996). *Motor-Free Visual Perception Test–Revised*. Novato, CA: Academic Therapy Publications.
- Colarusso, R. P., & Hammill, D. D. (2003). *Motor-Free Visual Perception Test–3*. Novato, CA: Academic Therapy Publications.
- Eliasson, A. C., Krumlinde-Sundholm, L., Rosblad, B., Beckung, E., Arner, M., Ohrvall, A. M., et al. (2006). The Manual Ability Classification System (MACS) for children with cerebral palsy: Scale development and evidence of validity and reliability. *Developmental Medicine and Child Neurology, 48*, 549–554.
- Flansbjerg, U. B., Holmback, A. M., Downham, D., Patten, C., & Lexell, J. (2005). Reliability of gait performance tests in men and women with hemiparesis after stroke. *Journal of Rehabilitation Medicine, 37*, 75–82.
- Gardner, M. F. (1996). *TVPS–R: Test of Visual–Perceptual Skills (Non-Motor)–Revised manual*. Hydesville, CA: Psychological and Educational Publications.
- Gliner, J. A., Morgan, G. A., & Harmon, R. J. (2001). Measurement reliability. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 486–488.
- Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy, 86*, 735–743.
- Ito, J., Saijo, H., Araki, A., Tanaka, H., Tasaki, T., Cho, K., et al. (1996). Assessment of visuo-perceptual disturbance in children with spastic diplegia using measurements of the lateral ventricles on cerebral MRI. *Developmental Medicine and Child Neurology, 38*, 496–502.
- Koeda, T., & Takeshita, K. (1992). Visuo-perceptual impairment and cerebral lesions in spastic diplegia with preterm birth. *Brain and Development, 14*, 239–244.
- Kozeis, N., Anogeianaki, A., Mitova, D. T., Anogianakis, G., Mitov, T., Felekidis, A., et al. (2006). Visual function and execution of microsaccades related to reading skills in cerebral palsied children. *International Journal of Neuroscience, 116*, 1347–1358.
- Kulp, M. T., Edwards, K. E., & Mitchell, G. L. (2002). Is visual memory predictive of below-average academic achievement in second through fourth graders? *Optometry and Vision Science, 79*, 431–434.
- Marozas, D. S., & May, D. C. (1986). Research on effects of color reversal on the visual perceptual and visuomotor performances of spastic cerebral palsied and other exceptional individuals. *Perceptual and Motor Skills, 62*, 595–607.
- McDonough, S. C., & Cohen, L. B. (1982). Attention and memory in cerebral palsied infants. *Infant Behavior and Development, 5*, 347–353.

- McFall, S. A., Deitz, J. C., & Crowe, T. K. (1993). Test-retest reliability of the Test of Visual Perceptual Skills with children with learning disabilities. *American Journal of Occupational Therapy, 47*, 819-824.
- Menken, C., Cermak, S. A., & Fisher, A. (1987). Evaluating the visual-perceptual skills of children with cerebral palsy. *American Journal of Occupational Therapy, 41*, 646-651.
- Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology, 39*, 214-223.
- Pirila, S., van der Meere, J., Korhonen, P., Ruusu-Niemi, P., Kyntaja, M., Nieminen, P., et al. (2004). A retrospective neurocognitive study in children with spastic diplegia. *Developmental Neuropsychology, 26*, 679-690.
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: An illustration of appropriate statistical analyses. *Clinical Rehabilitation, 12*, 187-199.
- Teller, D. (1990). *Teller Acuity Card handbook*. Dayton, OH: Vistech Consultants.