Commentary

Hong Kong Journal of Occupational Therapy

Hong Kong Journal of Occupational Therapy
2018, Vol. 31(2) 59–68

DOI: 10.1177/1569186118793847
journals.sagepub.com/home/hjo

An overview and critique of the *Test of Visual Perception Skills – fourth edition* (TVPS-4)

Ted Brown and Lisa Peres

Abstract

Introduction: The *Test of Visual Perception Skills – fourth edition* (TVPS-4) is an updated version of the *Test of Visual Perception Skills – third edition* (TVPS-3). The TVPS-4 is a standardized assessment of motor-free visual perception skills for individuals aged 5–21 years. Test norms were derived from a normative sample that reflected the United States population.

Method: The present paper outlines and appraises the reliability and validity of the TVPS-4. Measurement properties of the TVPS-4 were critiqued using the Quality Criteria for Health Status Questionnaires (QCHSQ) and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist with 4-point rating scale.

Results: The TVPS-4 manual reports details of reasonable levels and types of overall reliability and validity. The QCHSQ and COSMIN checklist reviews suggested that further research is needed in relation to the TVPS-4's measurement error, reproducibility, internal consistency and cross-cultural validity.

Conclusion: The TVPS-4 is an in-depth assessment that can be used to identify areas of impairment and assist with goal setting and intervention planning. The impact of the TVPS-4 on occupational therapy practice is also considered. Future research could investigate the cross-cultural validity of the TVPS-4 so that it can be used in other countries with confidence.

Keywords

Children, COSMIN, reliability, validity, visual perception

Received 22 January 2018; accepted 19 June 2018

The *Test of Visual Perception Skills – fourth edition* (TVPS-4; Martin, 2017) is a revised version of the *Test of Visual Perception Skills – third edition* (TVPS-3; Martin, 2006). The TVPS-4 assesses two-dimensional visual perception skills that do not require a motor response, such as copying (Martin, 2017). Test-takers are presented with a number of black and white visual stimuli on test plates and indicate their response verbally or through gestures. The TVPS-4 can be used by occupational therapists, educators, school psychologists, optometrists and other professionals who need to assess the visual perception abilities of individuals in childhood, adolescence or young adulthood ages 5 to 22 years of age (Martin, 2017). The results from the TVPS-4 are intended to be used for diagnosis, goal-setting, intervention planning and research (Martin, 2017).

The present paper aims to describe and critique the TVPS-4 using the Quality Criteria for Health Status Questionnaires (QCHSQ; Terwee et al., 2007) and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN; Mokkink et al., 2010; Terwee et al., 2012) checklist with a 4-point rating scale. This was done to provide prospective users of the TVPS-4 with a summary and appraisal of

Department of Occupational Therapy, Monash University, Australia

Corresponding author:
Ted Brown, Department of Occupational Therapy, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University – Peninsula Campus, Frankston, Victoria 3199, Australia.
Email: ted.brown@monash.edu

the scale's psychometric properties. A comparison of the TVPS-4 with two other well-known visual perceptual assessments (e.g., *Developmental Test of Visual Perception – third edition* (DTVP-3) and the *Motor-Free Visual Perception Test – fourth edition* (MVPT-4)) is also included.

Description of the test of visual perceptual skills-4 (TVPS-4)

The TVPS-4 'is an individually administered assessment of two-dimensional visual-perceptual skills for individuals age 5 through 21' (Martin, 2017, p. 7). There have been three previous iterations of the TVPS-4 referred to as the *Test of Visual-Perceptual Skills (non-motor)* (Gardner, 1982), *Test of Visual Perceptual Skills (non-motor) – revised* (Gardner, 1996) and *Test of Visual Perceptual Skills – third edition* (Martin, 2006).

To contextualize the domains of visual perception that are assessed by the TVPS-4, two recent theoretical models of visual perception are used: Scheiman's (2011) Model of Visual Information Processing (MoVIP) and the Cattell–Horn–Carroll Theory of Cognitive Abilities (CHCToCA) (Martin, 2017). There are three factors in Scheiman's (2011) MoVIP: visual integrity, visual efficiency and visual information processing. The visual information processing factor encompasses visual spatial, motor and analysis skills (Scheiman, 2011).

The TVPS-4 assesses the visual analysis skills that are tied to the visual information processing factor (Martin, 2017). Visual analysis skills refer to 'the ability of the child to be aware of the distinctive features of visual forms, including shape, size, color and orientation' (Scheiman, 2011, p. 80). The overall visual analysis skill has four subsets: *visual* discrimination, visual figure ground, visual memory and visual closure (Martin, 2017). Visual discrimination is the ability to distinguish forms based on their characteristics; visual figure ground involves perceiving a figure while filtering out an irrelevant background; visual memory is the ability 'to recognize and recall visually presented information' (Scheiman, 2011, p. 81); and visual closure is the ability to identify a complete image when a partially complete image is presented (Howe, Chen, Lee, Chen, & Wang, 2017).

The CHCToCA is a model that is supported by developmental, neurocognitive and outcome-criterion literature (Flanagan, 2013). In the CHCToCA model, cognitive abilities are categorized as either broad or narrow abilities (Flanagan, 2013). There are two types of broad cognitive abilities: domain-specific and domain-free (Flanagan & Harrison, 2012). Domain-specific cognitive abilities are linked to sensory factors whereas domain-free cognitive factors are not linked to sensory systems (Flanagan & Harrison, 2012; Pase & Stough, 2014). Visual processing is a domain-specific factor that consists of 11 narrow abilities including spatial relationships, spatial scanning and length estimation (Pase & Stough, 2014).

The TVPS-4 measures three of the 11 narrow abilities, namely visualization, flexibility of closure and visual memory (Martin, 2017). Visualization is thought to be encompassed within Scheiman's visual processing factor and flexibility of closure is similar to visual figure ground in Scheiman's MoVIP (Martin, 2017). Visual memory is believed to be more specific than the visual memory and visualization skills from Scheiman's model (Martin, 2017).

The TVPS-4 assesses visual perception skills via the completion of seven subtests and is therefore a good fit with the MoVIP and CHCToCA given that these models identify multiple domains of visual perception. The seven TVPS-4 subtests are *visual discrimination, visual memory, spatial relationships, form constancy, sequential memory, visual figure ground* and *visual closure* (Martin, 2017). The entire test, or a single subtest, can be administered to a participant and a raw score is obtained for each subtest by adding together the number of correctly answered items (Martin, 2017). Raw scores are then converted to standardized scores.

The TVPS-4 remains an individually administered test of visual perception ability and has retained the structure and multiple-choice response format of the TVPS-3. However, the TVPS-4 has been revised by the addition of two lower level items in each subtest. This was completed to make the TVPS-4 more suitable for younger prospective test-takers and those who have greater degrees of visual perception impairment. Norms in the TVPS-4 have been extended to age 21 to allow for the assessment of young adults.

A uniform start-point based on the test-taker's age was also added for persons aged 12 and over to reduce the testing burden on older respondents. Subtest differential scores, which were not available in the TVPS-3, were added to the TVPS-4. This scoring system allows test-users to determine whether there is a statistical difference between subtest scores. The TVPS-4 test manual has been updated based on current literature and the record form has also been updated to facilitate ease of scoring (Martin, 2017).

Norms

The TVPS-4 norms were derived from a sample of 1790 individuals aged between 5 years 0 months to 21 years 11 months that were representative of the United States (US) population. The demographics of the normative sample were compared to US population census data

from 2010 and closely matched that of the US population (Martin, 2017). Test-takers should be cautious when using the TVPS-4 to assess the visual perception of people who are not reflective of the US normative sample.

Reliability

To provide evidence for the TVPS-4's reliability, the test manual examined internal-consistency reliability, test–retest reliability, standard error of measurement (SEM) and confidence intervals (CI). Cronbach's alpha, a measure of internal-consistency reliability, was calculated for the TVPS-4's subtest scores and overall scores in each age group. Cronbach's alpha coefficients that range from 0.70 to 0.90 are considered acceptable (Tavakol & Dennick, 2011). On the individual subtest scores, the average alpha value was from 0.68 to 0.81. The average Cronbach's alpha value for the TVPS-4's overall scores was 0.94. The Cronbach's alpha for the TVPS-4's subtest scores in each age group ranged between 0.43 and 0.90, while the overall score in each age group ranged from 0.91 to 0.96. Most Cronbach's alpha values were above the 0.70 cut-off, providing support for the TVPS-4's internal-consistency reliability. However, 18 out of 105 TVPS-4 subtest Cronbach's alpha values at each age group were below 0.70.

The TVPS-4 demonstrated adequate test–retest reliability (range of 14–25 days between first and second test completion), with average corrected correlation coefficients ranging from 0.46 to 0.81 for subtest scores, and 0.97 for the overall score. The SEM is the 'standard deviation of errors that are associated with test scores from a particular group of examinees' (Harvill, 1991, p. 181). The SEM for TVPS-4 subtest scores (ranging from 0.97 to 2.26) and overall scores (ranging from 3.04 to 4.42) were reported by age with 90% CI and 95% CI, respectively (refer to Table 1). It is important to note that the SEM was calculated based on Cronbach's alpha values.

Validity

The TVPS-4 manual reported evidence for content validity, construct validity and criterion validity. Content validity is 'the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose' (Haynes, Richard, & Kubany, 1995, p. 238). To ensure good content validity, item difficulty and item discrimination were computed for each TVPS-4 subtest and the overall score. TVPS-4 subtest item difficulty by age ranged from 0.30 to 0.90, and overall item difficulty by age ranged from 0.36 to 0.83. Overall item difficulty increased with age, thus demonstrating that children found the TVPS-4 items more difficult than young adults. This follows the developmental trajectory expected for visual perceptual skills from childhood through to young adulthood (Martin, 2017).

There was adequate item discrimination; average item discrimination index by age for the TVPS-4 subtest scores ranged between 0.27 and 0.58, and the average overall item discrimination index by age ranged between 0.27 and 0.40 (Martin, 2017). The TVPS-4 items were assessed for item bias using differential item functioning (DIF) procedures. Group differences in the normative respondent sample that were investigated included: gender (male/female), residence (urban-suburban/rural), ethnicity (African American/non-African American; Caucasian/non-Caucasian; Asian American/non-Asian American) and Hispanic origin (Hispanic/non-Hispanic) (Martin, 2017). There were 10 instances of DIF occurring among nine items in the TVPS-4; however in eight instances the bias effect size was negligible (Martin, 2017).

The content validity of the TVPS-4 is also linked to its construct validity (Anastasi & Urbina, 1997). A summary of the TVPS-4's content validity is reported in Table 1. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) reporting a one factor solution for the TVPS-4 were presented as content validity evidence by Martin (2017). The EFA factor loadings for the seven TVPS-4 subscales ranged from 0.56 to 0.78. It appears that the TVPS-4's one factor solution is more closely related to its structural validity than content validity (Mokkink et al., 2010).

Construct validity is defined as 'the degree to which an assessment instrument measures the targeted construct' (Haynes et al., 1995, p. 239). To demonstrate construct validity, the authors tested a series of assumptions in relation to the TVPS-4. Firstly, the TVPS-4 scores were expected to be positively correlated with age, with a moderate to large correlation. It was found that in the normative sample, TVPS-4 raw scores were significantly positively correlated with age, thus supporting the idea that the TVPS-4 assesses skills that evolve with children's maturation and skill development (Martin, 2017).

Visual perception has been shown to be significantly worse among children with learning disabilities (Pieters, Desoete, Roeyers, Vanderswalmen, & Van Waelvelde, 2012). Individuals with autism spectrum disorder (ASD) also have a unique visual perception style (Behrmann, Thomas, & Humphreys, 2006; Vlamings, Jonkman, van Daalen, van der Gaag, & Kemner, 2010). Therefore, the TVPS-4 manual hypothesized that children diagnosed with a learning disability or ASD would have lower TVPS-4 scores than typically

Table 1. A summary of the reliability and validity for the *Test of Visual Perception Skills – fourth edition* (TVPS-4).

Reliability	Validity
Internal consistency (Cronbach's alpha)	Content validity
Subtest scores: average alpha ranged from 0.68 to 0.81 as reported in test manual.	Subtest item difficulty by age: from 0.30 to 0.90.
Overall scores: average of 0.94.	Overall item difficulty by age: from 0.36 to 0.83.
Each age group subtest scores: ranged from 0.43 to 0.90 as reported in the test manual ($n = 1790$).	Average item discrimination by age for subtest scores: item discrimination index ranged from 0.27 to 0.58.
Each age group overall scores: ranged from 0.91 to 0.96 as reported in the test manual.	Average overall item discrimination by age: 0.27–0.40.
Test-retest reliability ($n = 71$, 14–25 days)	Differential Item Functioning (DIF)
Subtest scores: corrected correlation coefficient ranged from 0.46 to 0.81 ($p < .001$).	Examined for: gender (male/female), residence (urban-suburban/rural), ethnicity (African American/non-African American; Caucasian/non-Caucasian; Asian American/non-Asian American) and Hispanic origin (Hispanic/non-Hispanic).
Overall score: corrected correlation coefficient $r = 0.97$ ($p < .001$).	Nine items had significant DIF. Eight instances of DIF had negligible bias effect size. There was DIF with moderate bias effect size between Hispanic and non-Hispanic children. There was DIF with large bias effect size between urban-suburban and rural children.
Standard Error of Measurement (SEM) and Confidence Intervals (CI) by age groups	Construct identification validity
SEM of subtest scores: ranged from 0.97 to 2.26 as reported in test manual.	Relationship to age: raw scores positively correlated with age (overall $r = 0.65$, $p \leq .001$).
SEM of overall score: ranged from 3.04 to 4.42 as reported in test manual.	Differences among groups: standard scores of groups with learning disabilities, Autism spectrum disorder and attention deficit hyperactivity disorder were compared with non-diagnosed individuals ($p < .001$).
90% CI of subtest scores: ranged from 1.60 to 3.73.	
95% CI of standard scores: ranged from 1.91 to 4.44.	
90% CI of overall score: ranged from 5.00 to 7.28.	
95% CI of overall score: ranged from 5.95 to 8.67.	
95% CI of overall score: ranged from 5.95 to 8.67.	
	Criterion validity (concurrent validity) TVPS-4 scores: the corrected correlation coefficient for the correlation between TVPS-4 scores and *Motor-free Visual Perception Test – fourth edition* scores were $r = 0.90$ ($p < 001$) ($n = 32$).

TVPS-4: *Test of Visual Perceptual Skills – fourth edition*; SEM: Standard Error of Measurement; CI: Confidence Interval.
Source: Martin, N. A. (2017).

developing children (Martin, 2017). People with attention deficit hyperactivity disorder (ADHD) without any other comorbid conditions (such as learning disabilities) do not have difficulties with visual perception; therefore, the TVPS-4 author assumed that TVPS-4 scores of children with ADHD would not be lower than non-diagnosed children (Martin, 2017).

Children, teenagers and young adults from the normative sample who were diagnosed with a learning disability, ASD or ADHD were matched with non-diagnosed participants on age, gender, ethnicity and Hispanic origin (Martin, 2017). Consistent with the predictions, participants with a learning disability or ASD had significantly lower scores on the TVPS-4 compared to the matched healthy participants (Martin, 2017). Those with an ADHD diagnosis had lower mean TVPS-4 scores than their matched non-diagnosed participants, but this was not statistically significant (Martin, 2017). Overall, these group differences provide evidence for the TVPS-4's construct validity. This is an important measurement property for an assessment tool that is used with clinical populations.

Criterion validity considers whether a test measures what it purports to measure by comparing it to an instrument that measures a similar skill, ability, attribute or factor (Field, 2013). The TVPS-4 was compared to the MVPT-4 (Colarusso & Hammill, 2015), a standardized test of visual perceptual skills. The TVPS-4 and MVPT-4 are both measures of visual perception and, therefore, it is plausible that they are associated (Colarusso & Hammill, 2015; Martin, 2017). In a

sample of 32 test-takers (aged 7 years to 16 years 11 months; 34% female, 66% male), the TVPS-4 and MVPT-4 scores were significantly correlated (r corrected = .90, $p < .001$) (Martin, 2017). Furthermore, the mean difference between the TVPS-4 and MVPT-4 scores was non-significant ($p = .253$) (refer Table 1) (Martin, 2017).

A comparison of the TVPS-4, DTVP-3 and the MVPT-4

The TVPS-4, DTVP-3 and MVPT-4 all assess motor-free visual perceptual skills. However the three scales are somewhat different in their scope, breadth, number of items, number of subscales, and measurement properties. These features are summarized in Table 2. The age range for the three scales varies: TVPS-4, 5 to 21 years of age; DTVPS-3, 4 to 12 years of age; and MVPT-4, 4 to 80+ years of age. The TVPS-4 assesses seven discreet types of motor-free visual perceptual skills while the DTVP-3 assesses three subtypes of motor-free visual perceptual skills. On the other hand, the MVPT-4 only generates one summary motor-free visual perceptual ability score. A unique feature of the DTVP-3 is that it also includes two visual-motor integration subscales and provides three composite scale scores: visual motor integration composite, motor-reduced visual perception composite and general visual perception composite. All three of the scales provide raw scores, scaled scores, percentile rank and age-equivalents.

The TVPS-4 manual provides limited information about its validity mainly based on EFA and CFA evidence, convergent validity and know-group difference validity. The MVPT-4's validity evidence is scant at best reporting information on content validity, convergent validity and DIF. The DTVP-4 has the most robust range of validity evidence reported including content-description validity, criterion-prediction validity, and construct-identification validity. All three scales report reasonable reliability evidence that includes internal consistency and test–retest reliability.

A critique of the TVPS-4

Several criteria can be applied by researchers to summarize and critically evaluate the measurement properties of standardized assessment tools. This review paper used the QCHSQ (Terwee et al., 2007) and the COSMIN (Mokkink et al., 2010; Terwee et al., 2012) to assess and critique the strengths and weaknesses of the TVPS-4 on a number of measurement properties.

Quality Criteria for Health Status Questionnaires

The purpose of the quality criteria developed by Terwee et al. (2007) was to evaluate the measurement properties of health status questionnaires. The QCHSQ provides standards for assessing the following: content validity, internal consistency, criterion validity, construct validity, reproducibility, responsiveness, floor and ceiling effects and interpretability (Terwee et al., 2007). Measurement properties can be given a rating of positive, negative or indeterminate and a study is rated indeterminate when there are methodological flaws in its design (Terwee et al., 2007). Table 3 summarizes the measurement properties of the TVPS-4, as suggested by Terwee et al. (2007). For further details on the QCHSQ, please refer to Terwee et al. (2007).

Two strengths of the TVPS-4 are that it has good construct validity and criterion validity. Criterion validity evidence comes from the comparison between the TVPS-4 and MVPT-4 scores. The TVPS-4 test manual does not explicitly argue that the MVPT-4 is a gold standard comparison measure, but instead states that it is reasonable to assume this given that the MVPT-4, like the TVPS-4, is also a measure of motor-free visual perception that was also developed and normed in the US (Colarusso & Hammill, 2015; Martin, 2017). However, the MVPT-4 is somewhat different from the TVPS-4 in that it does not contain specific visual perceptual subtests, but instead provides an overall score of visual perception ability (Colarusso & Hammill, 2015). The correlation between TVPS-4 scores and MVPT-4 score was 0.90 ($r > 0.70$), thus indicating good criterion validity (Martin, 2017; Terwee et al., 2007).

Construct validity was given a positive rating on the QCHSQ because specific hypotheses about differences in scores between groups were formulated a-priori (Terwee et al., 2007). The TVPS-4's author hypothesized that TVPS-4 scores would have a moderate to large positive correlation with chronological age. It was predicted that children with a diagnosed learning disability would have lower TVPS-4 scores than their non-diagnosed age and gender matched counterparts. Similarly, children diagnosed with ASD were also expected to have lower TVPS-4 scores than those without an ASD diagnosis. Participants with an ADHD diagnosis but no comorbid conditions were not assumed to have lower TVPS-4 scores than non-diagnosed peers. Indeed, 100% of the results supported the hypotheses, thus surpassing the QCHSQ's requirements of 75% (Martin, 2017; Terwee et al., 2007).

In order to provide further evidence of the TVPS-4's construct validity, an EFA was performed on the normative sample. This revealed that the TVPS-4 had one underlying factor (Martin, 2017). The EFA was then

Table 2. A comparison of the Test of Visual Perceptual Skills – 4 (TVPS-4), *Developmental Test of Visual Perception – third edition* (DTVP-3), and the Motor-free Visual Perception Test – 4 (MVPT-4).

	TVPS-4	DTVP-3	MVPT-4
Authors	Martin	Hammill, Pearson, and Voress	Colarusso and Hammill
Date of latest revision	2017	2014	2015
Publisher	ATP assessments	Pro-Ed	ATP assessments
Size & location of normative group	1790 children and adolescents from four geographical regions in the USA (e.g., North Central, Northeast, South, West)	1035 children from 27 states in the USA	2700 children and adults from four regions in the USA (e.g., North Central, North East, South, West) that included representation from 25 different states
Age range	5–0 to 21–11 years of age	4–0 to 12–11 years of age	4–0 to 80+ years of age
Types of visual perceptual skills assessed	• Visual discrimination • Visual memory • Spatial relationships • Form constancy • Sequential memory • Visual figure-ground • Visual closure	• Visual motor integration composite composed of eye-hand coordination & copying • Motor-reduced visual perception composite composed of three subscales: figure-group, visual closure & form constancy	• Spatial relationships • Visual discrimination • Figure-ground • Visual memory • Visual closure
Number of subscales & composite scales	Seven subscales: • Visual discrimination • Visual memory • Spatial relationships • Form constancy • Sequential memory • Visual figure-ground • Visual closure	Five subscales: • Eye-hand coordination • Copying • Figure-group • Visual closure • Form constancy Three composite scales: • Visual motor integration composite • Motor-reduced visual perception composite • General visual perception	One composite scale consisting of 45 items
Administration time	30 min	20–40 min	20–30 min
Time to score	5–10 min	15–20 min	5 min
Types of scores provided	• Raw scores • Scaled scores • Standard scores • Percentile ranks • Age equivalents • T-score • Stanine	• Raw scores • Scaled scores • Percentile ranks • Age equivalents	• Raw scores • Standard scores • Percentile ranks • Age equivalents • T-score • Stanine • Scaled scores
Types of validity evidence reported	• Exploratory factor analysis • Confirmatory factor analysis • Convergent validity	• Content-description validity: item discrimination & differential item functioning	• Item bias: differential item functioning • Validity based content • Comparisons with the *Test of Visual*

(continued)

Table 2. Continued

	TVPS-4	DTVP-3	MVPT-4
	• Know-group difference validity	• Criterion-prediction validity: correlation with criterion measures • Construct-identification validity: relationship to age, relationship amongst subscales and composite scales, differences amongst ages • Confirmatory factor analysis	*Perceptual Skills – 3rd edition* • Relationship to age
Types of reliability evidence reported	• Internal consistency – Cronbach alpha • Temporal stability (test–retest reliability)	• Internal consistency – Cronbach alpha • Test–retest reliability • Inter-scorer reliability (inter-rater reliability)	

Table 3. Measurement property ratings for the *Test of Visual Perception Skills – fourth edition* (TVPS-4) according to the Quality Criteria for Health Status Questionnaires (QCHSQ).

QCHSQ property	Rating
Internal consistency	−
Reproducibility – agreement	()
Reproducibility – reliability	−
Content validity	+
Construct validity	+
Criterion validity	+
Responsiveness	()
Floor or ceiling effect	()
Interpretability	?

QCHSQ: Quality Criteria for Health Status Questionnaires.
Ratings: +: positive rating; −: negative rating; ?: indeterminate rating/unknown; (): no information available.

followed-up with a CFA on the normative sample to confirm the one factor solution that was found through the EFA (Martin, 2017). The CFA findings again indicated that a one factor solution fitted the data (Martin, 2017). The factor analysis sample size was considered appropriate ($n = 1790$) based on the guidelines cited in Terwee et al. (2007).

Two psychometric weaknesses of the TVPS-4 that were identified through the application of the QCHSQ were its internal consistency and reproducibility (reliability). An overall average Cronbach's alpha was calculated separately for each subtest; however, internal consistency was inadequate because Cronbach's alpha was below the 0.70 cut-off ($\alpha = 0.68$) for one subtest (Terwee et al., 2007). Terwee et al. (2007) specify that for continuous measures, a suitable measure for reproducibility (reliability) is the intraclass correlation coefficient (ICC). Notably, the Pearson's correlation coefficient is considered inadequate (Terwee et al., 2007). The TVPS-4's author used Pearson's correlation coefficients to evidence test–retest reliability; therefore, the reproducibility (reliability) measurement property was given a negative rating (Martin, 2017).

Two other notable weaknesses of the TVPS-4 were the lack of information provided in its test manual in relation to responsiveness as well as floor and ceiling effects. This is a notable omission and it is recommended that these two issues be addressed in the next revision of the TVPS-4.

Consensus-based Standards for the Selection of Health Measurement Instruments

The COSMIN is a measurement tool that aims to provide researchers with criteria for assessing the measurement properties of health measurement instruments (Mokkink et al., 2010). Each box on the COSMIN represents one measurement property; for instance, box A refers to internal consistency and box B refers to reliability (Mokkink et al., 2010). The COSMIN checklist with 4-point rating scale allows users to obtain an overall methodological quality score for

Table 4. Measurement property ratings for the *Test of Visual Perception Skills – fourth edition* (TVPS-4) according to the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN).

COSMIN property	Rating
Reliability	
Internal consistency	Fair
Reliability	Fair
Measurement error	Poor
Validity	
Content validity (including face validity)	Excellent
Structural validity (construct validity)	Fair
Hypothesis testing (construct validity)	Poor
Cross-cultural validity (construct validity)	–
Criterion validity	Fair
Responsiveness	–

COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments.

each measurement property 'box' (Terwee et al., 2012). Items included in the measurement property boxes are rated as either 'excellent', 'good', 'fair' or 'poor' (Terwee et al., 2012).

The methodological quality score for a measurement property is equal to the worst item score on that property (Terwee et al., 2012). For example, if any items in box D (content validity) are rated 'poor' then the overall rating for content validity is 'poor'. Note that interpretability and generalizability do not have a formal scoring system but can be used to extract relevant data (Terwee et al., 2012). The COSMIN checklist with 4-point rating scale was used to critically evaluate the measurement properties of the TVPS-4 (Terwee et al., 2012). A summary of the measurement property ratings of the TVPS-4 using the COSMIN is provided in Table 4.

Content validity is a notable strength of the TVPS-4. Items on the TVPS-4 represent areas of visual perception identified in past developmental literature and theoretical frameworks (Martin, 2006; Scheiman, 2011). Furthermore, items are dependent on age; they are most difficult for very young children and easiest for young adults (Martin, 2006). Factor analysis of the TVPS-4 showed that the test is a measure of abilities that are encompassed within visual perception (Martin, 2006). The TVPS-4 has a one factor solution, such that items in each subtest combine to comprehensively reflect an overall comprehensive visual perception factor (Martin, 2006). Overall, this information provides evidence for the TVPS-4's excellent content validity.

Both measurement error and hypothesis testing measurement properties are shortcomings of the TVPS-4. These properties were both given a rating of 'poor'. The SEM, a measurement error statistic, was calculated using inappropriate statistical methods. The SEM that was presented in the TVPS-4 manual was calculated based on Cronbach's alpha, therefore constituting a low COSMIN 4-point rating (Martin, 2017).

The information relating directly to the hypotheses themselves in the hypothesis testing section was sufficient, but convergent validity (contained within the hypothesis testing measurement property of the COSMIN) was inadequate. The TVPS-4 was compared against the MVPT-4 to investigate convergent validity (Martin, 2017). However, the TVPS-4 description of the MVPT-4 lacked sufficient detail and the test manual did not outline the measurement properties of the MVPT-4 or cite studies that described these properties (Martin, 2017). Subsequent editions of the TVPS require a more comprehensive description of comparator instruments and their related measurement properties. Also, given a number of fundamental differences between the MVPT-4 and the TVPS-4 in relation to the number of items, subscale structure (or the lack thereof) and age range of the normative data, it is questionable whether the MVPT-4 was the most appropriate comparator scale for the TVPS-4.

Implications for occupational therapy practice

Visual perception deficits affect children and young adults' ability to engage in their daily occupations including work, education, self-care, play, leisure and social participation (Schneck, 2005). Therefore, it is essential that occupational therapists screen for, identify and address visual perception deficits to facilitate clients' return to optimal occupational performance. For example, previous research has identified that visual perception deficits are more common in children who have cerebral palsy (CP) (Barca, Cappelli, Di Giulio, Staccioli, & Castelli, 2010; Ego et al., 2015; James, Ziviani, Ware, & Boyd, 2015). A review by Ego et al. (2015) suggested that visual perception problems are a core feature of CP. Indeed, in children and adolescents with CP, specific visual perception problems with visual closure, visual sequential memory, visual memory and figure ground are associated with activities of daily living (James et al., 2015).

Screening tools (like the MVPT-4) can be used to determine whether visual perception problems exist. If visual perception problems are indeed identified, it requires a more comprehensive, follow-up test to be administered to pinpoint the domain of visual perception impacted (Cooke, McKenna, & Fleming, 2005). The TVPS-4 is an in-depth visual perception

assessment tool that occupational therapists can use to identify areas of visual perception impairment and evaluate intervention efficacy post-intervention. All seven subtests, each representative of distinct visual perception domains, can be administered to gain an overview of a child or young adult's visual perception abilities (Martin, 2017).

For example, if a young adult had a low subtest score on visual memory, it could be speculated that the functional implications of this could be difficulty remembering the items on their grocery shopping list, remembering phone numbers, and remembering where computer files are located and under what specific name. Unique to the present TVPS-4 version, examiners can also calculate subtest difference scores to gain insight into what areas test-takers found especially difficult (Martin, 2017). The visual perception information collected from the TVPS-4 can assist occupational therapists in generating treatment goals and choosing appropriate interventions that target and remediate specific deficits.

The TVPS-4 is a flexible, cost-effective and easily administered assessment of visual perception skills. Administration time of the full test is 30 min with a scoring time of between 5 and 10 min (Martin, 2017). The full TVPS-4 can be administered for a comprehensive assessment of visual perception. However, it is possible for test-users to administer a single subtest of the TVPS-4 to examine domains of visual perception. To reduce the testing burden on participants, basal and ceiling rules are included in the revised version of the scale (Martin, 2017). According to the QCHSQ and COSMIN checklist with 4-point rating scale, the TVPS-4 also demonstrates sufficient reliability and validity.

The primary limitation of the TVPS-4 is that it does not provide evidence for cross-cultural validity. The TVPS-4 was normed in a sample of children and adolescents (aged from 5 years 0 months to 21 years 11 months) that mirrored the US population based on demographic characteristics including gender, ethnicity, Hispanic origin and parents' education (Martin, 2017). Therefore, the TVPS-4 is best suited to test-takers from the US population. Further testing is needed to obtain normative data for populations outside the US. However, this requires time, money and resources. Until the cross-cultural validity of the TVPS-4 is established, occupational therapists who use the TVPS-4 in countries such as the United Kingdom, Australia, or other jurisdictions (such as Hong Kong or Taiwan) should exert caution when interpreting results based on US norms. There are also a lack of specific comparative TVPS-4 normative data for children presenting with different types of physical, intellectual, psychosocial, learning, and/or developmental disabilities. However, 149 children with a known disability (e.g., ASD ($n = 50$, 2.79%); attention deficit/hyperactivity disorder ($n = 43$, 2.40%), acquired brain injury ($n = 4$, 0.22%), CP ($n = 4$, 0.22%), specific learning disability ($n = 59$, 3.30%)) were included in the TVPS-4 normative sample. In other words, 8.32% of the total TVPS-4 normative sample had some type of known disability.

Another notable limitation of the TVPS-4 is the lack of reference norms for individuals older than 21 years of age. If occupational therapists wanted to use the TVPS-4 with adults or older adults, then the lack of comparator normative data to generate a profile of that group of test-takers visual perception skills is a significant limitation.

Conclusion

The TVPS-4 is a standardized measure of visual perception for children, adolescents and young adults aged from five to 21 years (Martin, 2017). It provides occupational therapists (and other education and clinical professionals) with a complete picture of an individual's visual perceptual skills. The information collected from the TVPS-4 can be used for goal setting, intervention planning and evaluation. Although the TVPS-4 has good overall reliability and validity, the QCHSQ and COSMIN checklist with 4-point rating scale suggest that further research is needed on certain measurement properties including measurement error, reproducibility (reliability) and internal consistency. Similarly, there is no evidence for the TVPS-4's cross-cultural validity. To ensure that occupational therapists interpret TVPS-4 scores appropriately in clinical practice, the TVPS-4 needs to be validated for use in other cross-cultural contexts. Further, if clinicians wish to use the TVPS-4 with adults older than 22 years of age, then additional normative data will need to be collected for this age group.

Authors' contributions

Dr Ted Brown and Ms Lisa Peres made equal contributions to reviewing the TVPS-4, analyzing the material and drafting the manuscript. Both contributing authors approved the final submitted manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed.). New York, NY: Macmillan.

Barca, L., Cappelli, F. R., Di Giulio, P., Staccioli, S., & Castelli, E. (2010) Outpatient assessment of neurovisual functions in children with cerebral palsy. *Research in Developmental Disabilities*, *31*(2), 488–495. DOI: 10.1016/j.ridd.2009.10.019

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences*, *10*(6), 258–264. http://dx.doi.org/10.1016/j.tics.2006.05.001

Colarusso, R., & Hammill, D. (2015). *Motor-free Visual Perception test-4 (MVPT-4)* (4th ed.). Novata, CA: Academic Therapy Publications.

Cooke, D. M., McKenna, K., & Fleming, J. (2005). Development of a standardized occupational therapy screening tool for visual perception in adults. *Scandinavian Journal of Occupational Therapy*, *12*(2), 59–71. https://doi.org/10.1080/11038120410020683-1

Ego, A., Lidzba, K., Brovedani, P., Belmonti, V., Gonzalez-Monge, S., Boudia, B., . . . Cans, C. (2015). Visual–perceptual impairment in children with cerebral palsy: A systematic review. *Developmental Medicine & Child Neurology*, *57*(Suppl 2), 46–51. DOI: 10.1111/dmcn.12687

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, UK: Sage Publications.

Flanagan, D. P. (2013). *The Cattell–Horn–Carroll Theory of Cognitive Abilities. In Encyclopaedia of special education*. Hoboken, NJ: John Wiley & Sons.

Flanagan, D. P., & Harrison, P. L. (2012) *Contemporary intellectual assessment: Theories, tests, and issues*. New York, NY: Guilford Press.

Gardner, M. F. (1982). *Test of Visual-Perceptual Skills (non-motor)*. San Francisco, CA: Psychological & Educational Publications.

Gardner, M. F. (1986). *Test of Visual-Motor Skills*. Seattle, WA: Special Child Publications.

Gardner, M. F. (1996). *Test of Visual Perceptual Skills (non-motor)—revised*. San Francisco, CA: Psychological & Educational Publications.

Hammill, D. D., Pearson, N. A., & Voress, J. K. (2014). *Developmental Test of Visual Perception – Third Edition: Examiners Manual*. Austin, TX: Pro-Ed.

Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practice*, *10*(2), 33–41.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. http://dx.doi.org/10.1037/1040-3590.7.3.238

Howe, T. H., Chen, H. L., Lee, C. C., Chen, Y. D., & Wang, T.N. (2017). The Computerized Perceptual Motor Skills Assessment: A new visual perceptual motor skills evaluation tool for children in early elementary grades. *Research in Developmental Disabilities*, *69*(Suppl C), 30–38. DOI?: 10.1016/j.ridd.2017.07.010

James, S., Ziviani, J., Ware, R. S., & Boyd, R. N. (2015). Relationships between activities of daily living, upper limb function, and visual perception in children and adolescents with unilateral Cerebral Palsy. *Developmental Medicine & Child Neurology*, *57*(9), 852–857. DOI: 10.1111/dmcn.12715

Martin, N. A. (2006). *Test of Visual Perceptual Skills* (3rd ed.). Novato, CA: Academic Therapy Publications.

Martin, N. A. (2017). *Test of Visual Perceptual Skills* (4th ed.). Novato, CA: Academic Therapy Publications.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. DOI: 10.1007/s11136-010-9606-8

Pase, M. P., & Stough, C. (2014). An evidence-based method for examining and reporting cognitive processes in nutrition research. *Nutrition Research Reviews*, *27*(2), 232–241. DOI: 10.1017/S0954422414000158

Pieters, S., Desoete, A., Roeyers, H., Vanderswalmen, R., & Van Waelvelde, H. (2012). Behind mathematical learning disabilities: What about visual perception and motor skills? *Learning and Individual Differences*, *22*(4), 498–504. https://doi.org/10.1016/j.lindif.2012.03.014

Scheiman, M. (2011). *Understanding and managing vision deficits: a guide for occupational therapists* (3rd ed.). Thorofare, NJ: Slack.

Schneck, C. M. (2005). Visual perception. In: J. Case-Smith (Ed.), *Occupational therapy for children* (pp. 412–448). St. Louis, MO: Elsevier Mosby.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. DOI: 10.5116/ijme.4dfb.8dfd

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657. DOI: 10.1007/s11136-011-9960-1

Vlamings, P. H. J. M., Jonkman, L. M., van Daalen, E., van der Gaag, R. J., & Kemner, C. (2010). Basic abnormalities in visual processing affect face processing at an early age in autism spectrum disorder. *Biological Psychiatry*, *68*(12), 1107–1113. DOI: 10.1016/j.biopsych.2010.06.024