The visual perceptual skills of children are often evaluated by health care and education practitioners. Even though the *Test of Visual Perceptual Skills – Revised* (TVPS-R) is one of the most frequently used instruments with school-age children, its construct validity has not been evaluated thoroughly. The purpose of the study was to evaluate the scalability/interval level measurement, unidimensionality, lack of differential item functioning (DIF), and hierarchical ordering of items of the TVPS-R and its seven subscales using the Rasch Measurement Model (RMM). The TVPS-R scores from a sample of 356 normally developing children (171 boys and 185 girls), ranging in age from 5 to 11 years, were used to complete the RMM analysis.

When the seven individual TVPS-R scales were analysed, they all exhibited adequate measurement properties (scalability/interval level measurement, unidimensionality, lack of DIF, and hierarchical ordering). However, when they were collapsed together to form an overall composite scale of motor-free visual perceptual skills, the TVPS-R items failed to group together to measure a unidimensional construct. In addition, many scale items exhibited RMM misfit or DIF.

The results suggest that the seven TVPS-R subscales can be used on an individual basis with clients to generate a profile of their motor-free visual perceptual skills, but that they cannot be summed together to calculate an overall summary motor-free visual perceptual score or perceptual quotient. The TVPS-R composite scale does not exhibit adequate construct validity.

An Evaluation of the Validity of the *Test of Visual Perceptual Skills – Revised* (TVPS-R) using the Rasch Measurement Model

Ted Brown[1] and Sylvia Rodger[2]

Introduction

Visual perception is an important area in occupational therapy (Daniels and Wong 1993, Gentile 1997, Scheiman 1997, Grieve 2000) and several tests of visual perception frequently used by occupational therapists include the *Motor-Free Test of Visual Perception – Revised* (Colarusso and Hammill 1996), the *Developmental Test of Visual Perception – 2* (Hammill et al 1993) and the *Test of Visual Perceptual Skills (Non-Motor) – Revised* (Gardner 1996). Practitioners need well-constructed, reliable and valid visual perceptual assessment tools that can be used with confidence. This study addressed this issue by evaluating the measurement properties of the *Test of Visual Perceptual Skills (Non-Motor) – Revised* (TVPS-R, Gardner 1996) with the Rasch Measurement Model (RMM, Bond and Fox 2007), a type of Item Response Theory model.

[1]Monash University – Peninsula Campus, Frankston, Victoria, Australia.
[2]The University of Queensland, Brisbane, Queensland, Australia.

Corresponding author: Dr Ted Brown, Senior Lecturer, Department of Occupational Therapy, School of Primary Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University – Peninsula Campus, Building G, 4th Floor, McMahons Road, PO Box 527, Frankston, Victoria 3163, Australia. Email: ted.brown@med.monash.edu.au

Submitted: 15 February 2008. *Accepted:* 7 January 2009.
Key words: Visual perception, testing, assessment, validity, children.
Reference: Brown T, Rodger S (2009) An evaluation of the validity of the Test of Visual Perceptual Skills – Revised (TVPS-R) using the Rasch Measurement Model. *British Journal of Occupational Therapy, 72(2),* 65-78.

Literature review

Visual perception

The completion of many educational activities and activities of daily living requires a combination of refined abilities, which includes vision, visual perception and visual motor functions (Chaikin and Downing-Baum 1997, Gentile 1997, Erhardt and Duckman 2005). Visual perception is understood here to include a person's ability to interpret, understand and define incoming visual information (Werner and Rini 1976, Scheiman 1997). Therefore, visual perception plays an important role in a person's daily functioning and occupational performance on many levels (American Occupational Therapy Association [AOTA] 1997, 2002, Kovacs 2000, Dankert et al 2003, Loikith 2005).

Optometrists, occupational therapists, psychologists and educators often assess and treat visual perceptual problems occurring in school-age children (Todd 1993, Weil and Amundson 1994, Fischer et al 2000, Wright et al 2000). Difficulties in this skill area can have a negative impact on a number of occupational performance and functional skill areas for children, which include problems in reading, spelling, cursive and manuscript written output, visual-motor integration, and mathematics (Solan and Ciner 1989, Schneck and Lemer 1993). In other words, visual perceptual dysfunction can affect negatively the ability of school-aged children successfully to complete their activities of daily living, participate in play or recreational activities, and complete school work (Groffman and Solan 1994, Parush et al 1998, Van Waelvelde et al 2004). This, in turn, can have a negative effect on their self-esteem and self-concept, and the accomplishment of age-related developmental tasks (AOTA 1991, 1994, Schneck 2001).

It is important, therefore, for occupational therapists to use assessments that possess sound measurement properties (for example, validity, reliability, responsiveness and clinical utility) in order to assess the presence and impact of visual perceptual dysfunction in children. One of the most commonly used tests of visual perceptual skills in school-age children is the *Test of Visual Perceptual Skills – Revised* (TVPS-R) (Hung et al 1987, Crowe 1989, Rodger 1994, Chia 1997, Chu and Chia 1997, Feder et al 2000, Bishop and Curtin 2001, Brown et al 2003, 2005, Brown and Gaboury 2006).

The Test of Visual Perceptual Skills – Revised (TVPS-R)

The TVPS-R is non-linguistically oriented and, according to Gardner (1982), the test developer, its stimuli are not culturally bound. It evaluates seven visual perceptual subskills: (1) visual discrimination, (2) visual memory, (3) visual spatial relationships, (4) visual form constancy, (5) visual sequential memory, (6) visual figure ground and (7) visual closure (Gardner 1982). There are 16 items, arranged progressively according to their difficulty, on each of the seven subscales. The subscales consist of various forms and shapes.

The TVPS-R is designed to be used with school-age children, between 4 and 12 years of age, who have four or five potential response options to choose from on each subscale item (Gardner 1982). They respond by selecting the correct choice from a multiple-choice format that does not require motor responses, such as drawing or copying shapes and designs. It is for this reason that the TVPS-R is referred to as being motor-free or non-motor. Depending on the age of the child, it takes approximately 30 to 45 minutes to administer the test and 5 to 10 minutes to score. The child is shown the test plates and asked to point to the correct response from among four or five choices on the card. Scoring is a simple matter of summing the correct responses on each subscale and determining derived scores. The subscale items are scored dichotomously. Limited validity and reliability data are reported in the TVPS-R manual (Brown et al 2003).

The Rasch Measurement Model (RMM)

The RMM, a type of Item Response Theory, can be used to evaluate the measurement properties of existing ordinal level instruments, in which items are intended to be summed together to provide a total score (Tennant and Conaghan 2007). Within the health science and education sectors, the RMM is progressively becoming one of the preferred methods of evaluating the construct validity of instruments, during construction, validation and evaluation (Wright et al 1993, Velozo et al 1999, Smith and Smith 2004, Bond and Fox 2007). With the RMM, it is possible to determine whether instruments possess interval level scaling, are unidimensional, possess stable item difficulty across different groups of participants and have items that are ordered from least to most difficult (hierarchical ordering) (Richardson 2005).

Interval level scaling requires that the units of measure reflect equal quantities across the range of the construct. Raw item scores, such as those obtained on the TVPS-R, are nominal and must be linearised so that they correspond to equal increments of the underlying construct. Goodness-of-fit statistics, expressed as mean square (MNSQ) infit/outfit statistics and standardised values, are used to determine how well the data from the items and participants fit the expectations of the RMM and thus reflect the construct being measured (Fischer and Molenaar 1995). A commonly accepted range for mean square values is 0.6 to1.4 and -2 to +2 for the standardised values. For fit to be out of range, both mean square and standardised values must be out of range (Bond and Fox 2007). Goodness-of-fit statistics found to be greater than the expected range indicate less predictable responses and suggest that the data may not fit the construct intended to be measured by the instrument. Fit statistics found to be lower than the RMM acceptable range suggest that the whole range of the scale may not be used; however, this is not considered as great a problem as values greater than the RMM range (Velozo et al 1999). In the present study, as the instrument being examined was already in print, a more stringent range of fit of 0.80 to 1.2 was used.

Unidimensionality indicates that the items of an instrument represent a single underlying dimension or construct, as evaluated by how the items fit the mathematical model (Wright and Stone 1979, Karabatos 2001). Goodness-of-fit statistics demonstrate the extent to which test items represent the single construct being measured; in this study, visual perceptual skills. Unidimensionality is confirmed if the instrument displays fit statistics within the RMM acceptable range, thus fitting the RMM requirements.

Differential item functioning (DIF) is the difference in the scoring of test items, based on gender, age or other variables, which occurs when participants respond differently to individual test items measuring equal levels of the underlying construct (Tennant and Conaghan 2007). DIF can affect the instrument's fit to the RMM if test items do not contain stable item difficulty across different groups of participants. In this study, DIF of the TVPS-R was established, based on gender, by comparing the person-ability logit scores of boys and girls on each of the TVPS-R test items.

Hierarchical ordering requires that items on an instrument be arranged from easy to difficult, representing the degree of the construct being measured (Hart et al 1997, Richardson 2005). Within the RMM, logit values represent the degree of difficulty of the test items, with the test items being ordered hierarchically from the easiest item to the most difficult item. Therefore, it is anticipated that item one of an instrument will be the easiest item, with subsequent test items consecutively increasing in difficulty. The last test item would be the most difficult.

Purpose

The purpose of this research was to examine the scalability and validity of the TVPS-R and its seven subscales, using the RMM methodology, with a sample of typical school-age children. It was hypothesised that:

1. The scalability/interval level scaling of the TVPS-R would be confirmed
2. The unidimensionality of the TVPS-R would also be confirmed
3. The differential item functioning of the item calibrations of the TVPS-R would be reproducible (stable) (for example, person-separation reliability) across different groups of participants, such as gender (for example, boys versus girls)
4. The TVPS-R and its respective subscales would each form hierarchical indexes with adequate item spacing, progressing from easier items to more challenging changing ones.

Method

Study design

The design was a prospective cross-sectional evaluation.

Participants

A sample of convenience of 356 children, ranging in age from 5 to 11 years, was recruited; of those, 171 or 48% were boys and 185 or 52% were girls. The participants were from one geographical area, that of the Ottawa metropolitan region, Ontario, Canada. Boys and girls were eligible for this study if (1) there was consent to participate in the study (by both the paediatric participant and his or her parent/guardian/caregiver); (2) they were between 5 and 11 years old; (3) they had proficient English speaking and listening skills; and (4) it was in the absence of any major diagnosed intellectual or physical impairment(s) by screening procedures.

A screening questionnaire was completed by each of the parents to ensure that the children met the inclusion criteria. The screening form included asked the parents if their child had any history of learning disabilities, had any medical diagnoses, had ever received extra assistance at school or had ever been referred to a health professional (such as a speech and language therapist or an occupational therapist). The rationale for excluding children diagnosed with intellectual or physical impairments from the study sample group was that the norms and developmental ordering of the scale items of the four instruments were based on the performance scores of a group of American children presenting with no intellectual or physical disability.

The children came from junior kindergarten through to grade 7. The total sample percentage distribution of the children in each grade level was as follows: junior kindergarten 3.1%, senior kindergarten 14.9%, grade one 16%, grade two 13.8%, grade three 16.3%, grade four 15.7%, grade five 9.3%, grade six 8.4%, and grade seven 2.5%. In the provincial education system in Ontario, Canada, junior kindergarten is designed for 4-year-old children and senior kindergarten for 5-year-old children. Half of the children were enrolled in the public school system (n = 178), 26.7% were enrolled in the Catholic school system (n = 95) and the remainder were enrolled in the private school system (23.3%). In the Ontario education system, both the public and Catholic school systems are publicly funded by the provincial government. The majority of the children spoke only English (71.3%), while the rest spoke English and French (25.6%), English and another language (1.7%), or English, French, and another language (1.4%).

Instrumentation

A demographic questionnaire was next used to gather relevant background data about the children. The children then completed the seven subscales of the TVPS-R. Details of the TVPS-R are located in the literature review.

Data analysis

The Statistical Package for the Social Sciences Version 10.0 (SPSS) (Kirkpatrick and Feeney 2001) was used for the data analysis. Descriptive statistics were calculated as appropriate to the data using SPSS.

Since the data set generated was dichotomous in nature, the RMM computer program, Winsteps, was used for the data analysis (Wright and Masters 1982, Smith 1991, 1992, Linacre and Wright 1998). RMM analysis is an iterative process, with the objective of achieving the 'best fit' of the data to the model by testing the model's assumptions. The objectives of the RMM analysis were to determine:

1. The scalability/interval level of measurement of the TVPS-R and its subscale items based on fit with the RMM
2. The unidimensionality of the TVPS-R and its subscales based on goodness-of-fit analysis

3. Whether differential item functioning (DIF) of the item calibration estimates occurred across participant samples in terms of gender
4. The hierarchical order and spacing of the TVPS-R and its subscale items based on item calibration (item difficulty parameter estimate) and standard error estimates.

Rasch Measurement Model analysis procedures

Item fit and confirmation of unidimensionality: The RMM evaluates the fit of the data to an unconditional probabilistic model. The logit values represent the difficulty of the items (item weights) in an instrument. With them, items are ordered from easiest to most difficult; this provides evidence of a hierarchical ordering of scale items. The fit of the items to the RMM was determined by the infit mean square statistic and the outfit mean square statistic, both of which are based on a chi-square distribution (Smith 1992). Fit statistics should range between 0.80 and 1.20. High or low fit statistics represent abnormalities in the response pattern to the item that may be related to a lack of unidimensionality, differential item functioning, poorly placed items in terms of developmental sequencing or poorly worded items (Linacre and Wright 1998). This step indicates how the items fit the RMM and provides information about an instrument's scalability and unidimensionality.

The infit and outfit statistics use slightly different methods for assessing an item's fit to the RMM. The infit statistic gives more weight to the performance scores of participants closer to the item value. The belief is that people whose ability is close to the item's difficulty will give a more sensitive insight into that item's performance (Bond and Fox 2007). The outfit statistic is not weighted and, therefore, is more sensitive to the influence of outlying scores: 'It is for this reason that users of the Rasch model routinely pay more attention to infit scores than outfit scores. Aberrant infit scores usually cause more concern than large outfit statistics' (Bond and Fox 2001, p43).

Differential item functioning (DIF): DIF, as evaluated by comparing the 95% confidence intervals of the logit values of the scale items, is based on gender. This required analysing the data set as a whole group and then in subgroups according to gender. This process of analysis confirmed the fit of the data to the RMM. As a result, the data were considered to be an interval level of measurement. This procedure mapped items onto the continuum of the latent trait, if one did exist. The average item calibrations from the RMM analysis defined the hierarchical order of the items along the continuum. Harder items were located at one end of the linear continuum and easier items were located at the opposite end.

Procedures

Ethics committee approval from the University of Queensland Behavioural and Social Sciences Ethical Review Committee, Brisbane, Queensland, Australia, and from the Children's Hospital of Eastern Ontario Ethical Review Committee, Ottawa, Ontario, Canada, were obtained. If informed consent was received from both the child and his or her parent/guardian/caregiver, the child was asked to complete the TVPS-R. The TVPS-R was administered by an occupational therapist who had 10 years of professional experience in administering the TVPS-R to children.

Since the purpose of the study was to evaluate the measurement properties of the TVPS-R, it was administered to each child in its entirety during one session instead of being discontinued when the child's performance reached the ceiling score outlined in the test manual. Under normal circumstances, when a child answers three consecutive questions wrongly or three out of four consecutive answers incorrectly on the subscales, his or her performance on that subscale is terminated. However, it was necessary to modify these standard instructions in order to evaluate the TVPS-R using the RMM.

Results

TVPS-R raw scores

The mean TVPS-R subscale scores for each age level are reported in Table 1.

TVPS-R Visual Discrimination Scale results

The TVPS-R Visual Discrimination (VD) scale consists of 16 dichotomously scored items. The 16 TVPS-R VD scale items were calibrated using the RMM and the results are reported in Table 2. All of the TVPS-R VD scale items were found to have infit and outfit statistics within acceptable parameters specified by the RMM. The TVPS-R VD scale items had mean square (MNSQ) infit statistics ranging from 0.82 to 1.28 and outfit statistics ranging from 0.49 to 1.29. The mean VD item measure was 0.00 logits (SD = 1.20). There was a moderately broad range of logit measures, the lowest value being -2.94 and the highest value being +2.23 (see Table 2). This resulted in an item separation index of 6.27 and a reliability of 0.98. The TVPS-R VD scale person-item map is located in Fig. 1. The average person measure was 1.60 logits (SD = 1.34). Person separation was 1.27, with a reliability of 0.62.

When the items were examined for DIF based on gender, they all fell within the 95% confidence interval. In other words, none of the TVPS-R VD scale items exhibited DIF based on gender. VD scale items 3 and 10 had low outfit scores, but were retained since their infit scores were in the 0.80 to 1.20 range. The final measure order of the 16 TVPS-R VD scale items (listed from easiest to most difficult) was as follows: 1, 5, 2, 3, 7, 15, 6, 8, 4, 12, 11, 10, 13, 9, 14 and 16. The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the TVPS-R VD scale.

TVPS-R Visual Memory Scale results

When the TVPS-R Visual Memory (VM) scale was calibrated using the RMM, VM scale item 12 was found to have infit and outfit statistics outside acceptable RMM parameters (see

Table 1. Mean TVPS-R scale scores based on age of participants (n = 356)

Age group		TVPS-R VD scale (range 0-16)	TVPS-R VM scale (range 0-16)	TVPS-R VSR scale (range 0-16)	TVPS-R VFC scale (range 0-16)	TVPS-R VSM scale (range 0-16)	TVPS-R VFG scale (range 0-16)	TVPS-R VC scale (range 0-16)	TVPS-R TS score (range 0-112)
5 years	Mean	8.86	8.56	9.98	8.61	7.44	9.96	7.12	60.54
(n = 57)	SD	3.36	2.86	3.57	2.68	2.67	2.52	2.82	15.58
6 years	Mean	10.80	10.39	11.93	10.05	9.14	10.71	8.96	72.00
(n = 56)	SD	3.22	2.58	3.00	2.94	3.44	2.77	2.92	16.29
7 years	Mean	12.86	11.45	14.00	11.52	11.13	12.73	10.84	84.52
(n = 56)	SD	2.34	2.46	2.17	2.43	2.70	2.14	2.87	12.76
8 years	Mean	13.53	12.49	14.65	12.09	12.05	13.67	12.33	90.81
(n = 57)	SD	1.79	2.00	1.56	2.59	2.38	1.62	2.76	10.68
9 years	Mean	14.33	13.71	15.33	13.29	13.35	14.55	13.82	98.36
(n = 55)	SD	2.18	1.37	.94	1.98	1.77	1.41	2.36	7.91
10 years	Mean	14.44	13.69	15.19	13.56	13.42	14.50	13.83	98.64
(n = 36)	SD	2.08	1.55	1.09	1.89	1.79	1.93	1.98	8.97
11 years	Mean	15.33	14.21	15.62	14.56	13.77	15.38	15.18	104.05
(n = 39)	SD	1.01	1.70	0.59	1.54	1.81	0.96	1.05	5.46

VD = Visual Discrimination; VM = Visual Memory; VSR = Visual Spatial Relationships; VFC = Visual Form Constancy; VSM = Visual Sequential Memory; VFG = Visual Figure Ground; VC = Visual Closure; and TS = Total Scale.

Table 3). All of the remaining TVPS-R items were found to have acceptable MNSQ infit and outfit statistics scores. Then, the TVPS-R VM scale items were examined for DIF based on gender. VM scale item 5 was found to exhibit DIF. Therefore, only 14 TVPS-R VM scale items fell within the acceptable RMM infit and outfit statistic parameter ranges.

The 14 TVPS-R VM scale items (excluding items 5 and 12) were found to have MNSQ infit statistics ranging from 0.85 to 1.20 and outfit statistics ranging from 0.50 to 1.33 (see Table 3). Item 1 had a low outfit score, but was retained since its infit score was in the 0.80 to 1.20 range. The mean VM item measure was 0.00 logits (SD = 1.67). There was a broad range of the logit measures, the lowest value being -3.82 and the highest value being +3.40. This resulted in an item separation index of 7.91 and a reliability of 0.98. The TVPS-R VM scale person-item map is located in Fig. 2. The average person measure was 1.59 logits (SD = 1.37). Person separation was 1.34 with a reliability of 0.64.

To reiterate, then, TVPS-R VM scale item 12 was discarded due to RMM misfit and VM item 5 was discarded due to DIF based on gender. The remaining 14 TVPS-R VM scale items were retained. The final measure order of these 14 VM scale items (listed from easiest to most difficult) was as follows: 1, 9, 3, 4, 2, 7, 10, 6, 8, 11, 14, 13, 15 and 16. The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the 14-item version of the TVPS-R VM scale.

Table 2. TVPS-R Visual Discrimination Scale item measure order (n = 356)

TVD STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVD
16	115	290	2.23	.14	1.15	1.8	1.10	.6	.22	TVD16
14	165	290	1.30	.14	1.01	.1	1.00	0.0	.35	TVD14
9	166	290	1.28	.14	1.04	.4	1.02	.1	.34	TVD9
13	176	290	1.09	.14	1.02	.2	.98	-.2	.36	TVD13
10	181	290	.99	.14	.82	-2.1	.69	-2.7	.52	TVD10
11	208	290	.41	.15	.94	-.6	.93	-.4	.45	TVD11
12	218	290	.17	.16	1.23	1.7	1.29	1.5	.23	TVD12
4	223	290	.05	.16	1.03	.2	.96	-.2	.38	TVD4
8	232	290	-.20	.17	.87	-.9	.80	-.9	.50	TVD8
6	234	290	-.26	.17	1.09	.6	1.07	.3	.33	TVD6
15	239	290	-.42	.18	.82	-1.2	.78	-.9	.53	TVD15
7	247	290	-.69	.19	1.28	1.4	1.17	.5	.20	TVD7
3	253	290	-.92	.20	.82	-1.0	.49	-1.9	.54	TVD3
2	256	290	-1.05	.21	.98	-.1	.88	-.4	.41	TVD2
5	256	290	-1.05	.21	.99	0.0	.89	-.3	.41	TVD5
1	282	290	-2.94	.38	1.05	.1	1.21	.2	.19	TVD1
MEAN	216.	290.	0.00	.18	1.01	0.0	.95	-.3		
S.D.	43.	0.	1.20	.06	.13	1.0	.20	1.0		

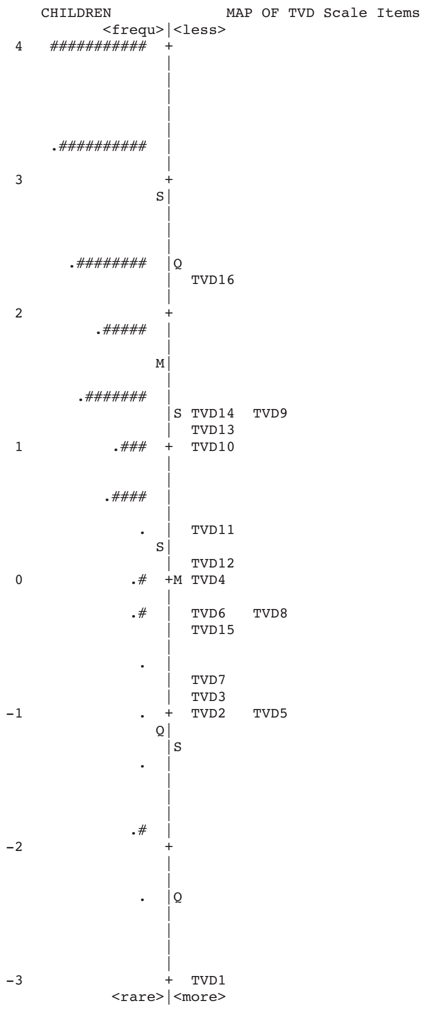
TVD = TVPS-R Visual Discrimination Scale; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

TVPS-R Visual Spatial Relationships Scale results

The TVPS-R Visual Spatial Relationships (VSR) scale items were calibrated using the RMM. VSR scale item 10 was found to have RMM misfit, whereas all of the remaining TVPS-R items were found to have acceptable RMM infit and outfit statistics scores (see Table 4). The items were also examined for DIF based on gender and VSR scale item 13 was found to exhibit DIF.

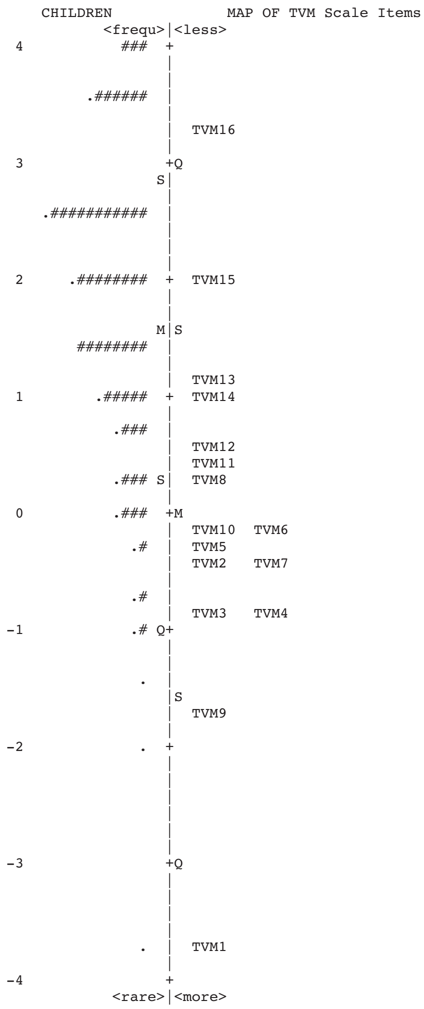
The 14 remaining TVPS-R VSR scale items were found to have MNSQ infit statistics ranging from 0.88 to 1.07 and outfit statistics ranging from 0.58 to 1.87. The mean VSR item measure was 0.00 logits (SD = 1.31). There was a limited range of logit measures, the lowest value being -2.46 and the highest value being +1.60. This resulted in an item separation index of 5.61 and a reliability of 0.97.

Fig. 1. TVPS-R Visual Discrimination Scale Rasch analysis person-item map (n = 356).



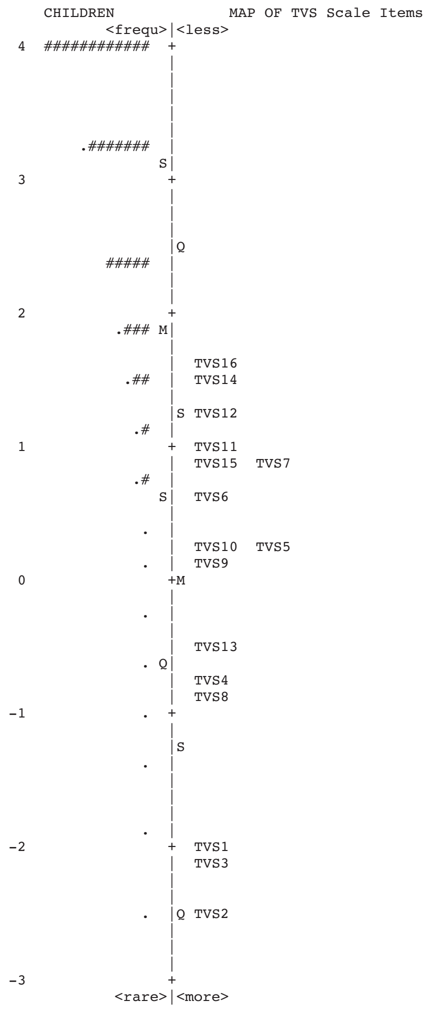
TVD = TVPS-R Visual Discrimination Scale item.

Fig. 2. TVPS-R Visual Memory Scale Rasch analysis person-item map (n = 356).



TVM = TVPS-R Visual Memory Scale item.

Fig. 3. TVPS-R Visual Spatial Relationships Scale Rasch analysis person-item map (n = 356).



TVS = TVPS-R Visual Spatial Relationships Scale item.

Table 3. TVPS-R Visual Memory Scale item measure order (n = 356)

TVM STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVM
16	74	338	3.23	.15	1.09	.8	1.21	.8	.15	TVM16
15	141	338	2.04	.13	.94	–.8	.91	–.7	.35	TVM15
13	194	338	1.20	.13	.98	–.3	.96	–.4	.34	TVM13
14	210	338	.94	.13	1.00	0.0	1.01	.1	.34	TVM14
12	233	338	.54	.14	**1.28**	**2.6**	**1.46**	**2.8**	**.12**	**TVM12**
11	239	338	.42	.14	.83	–1.7	.76	–1.7	.50	TVM11
8	245	338	.31	.14	.93	–.7	.98	–.1	.41	TVM8
6	264	338	–.08	.15	1.16	1.3	1.19	.9	.20	TVM6
10	267	338	–.15	.15	.85	–1.3	.88	–.6	.47	TVM10
5	275	338	–.34	.16	.98	–.2	.95	–.2	.35	TVM5
7	278	338	–.42	.16	.91	–.7	.75	–1.1	.41	TVM7
2	279	338	–.44	.16	.96	–.3	.78	–1.0	.37	TVM2
4	293	338	–.84	.18	1.09	.5	.94	–.2	.25	TVM4
3	294	338	–.87	.18	.92	–.5	.95	–.2	.37	TVM3
9	316	338	–1.78	.24	1.10	.4	1.18	.4	.15	TVM9
1	334	338	–3.74	.52	1.08	.1	.56	–.4	.15	TVM1
MEAN	246.	338.	0.00	.18	1.01	0.0	.97	–.1		
S.D.	64.	0.	1.51	.09	.12	1.0	.21	1.0		

TVM = TVPS-R Visual Memory Scale; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

RMM misfitting item: 12.

Table 4. TVPS-R Visual Spatial Relationships Scale item measure order (n = 356)

TVSS STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVSS
16	133	236	1.65	.15	1.08	1.1	1.21	1.6	.27	TVS16
14	141	236	1.46	.15	1.05	.7	1.03	.2	.34	TVS14
12	149	236	1.27	.16	.98	–.3	.92	–.7	.39	TVS12
11	159	236	1.01	.16	1.10	1.3	1.12	.8	.31	TVS11
15	162	236	.93	.16	1.00	0.0	1.03	.2	.39	TVS15
7	164	236	.88	.17	.99	–.1	.93	–.5	.40	TVS7
6	172	236	.65	.17	1.02	.2	.96	–.2	.39	TVS6
5	183	236	.31	.18	1.05	.5	1.21	1.0	.36	TVS5
10	183	236	.31	.18	**.72**	**–3.1**	**.54**	**–2.7**	**.63**	**TVS10**
9	188	236	.13	.19	1.10	.9	1.15	.7	.32	TVS9
13	203	236	–.47	.22	.85	–1.3	.86	–.5	.50	TVS13
4	208	236	–.72	.23	1.07	.5	1.14	.4	.32	TVS4
8	210	236	–.83	.24	.96	–.3	.59	–1.3	.43	TVS8
1	225	236	–1.97	.33	1.05	.2	1.34	.4	.25	TVS1
3	226	236	–2.09	.35	.90	–.4	1.16	.2	.32	TVS3
2	229	236	–2.51	.41	.96	–.1	2.15	1.0	.19	TVS2
MEAN	183.	236.	0.00	.22	.99	0.0	1.08	0.0		
S.D.	30.	0.	1.27	.08	.10	1.0	.35	1.0		

TVS = TVPS-R Visual Spatial Relationships Scale item; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

RMM misfitting item: 10.

The TVPS-R VSR scale person-item map is located in Fig. 3. The average person measure was 1.84 logits (SD = 1.22). Person separation was 0.78, with a reliability of 0.38.

In summary, TVPS-R VSR scale item 10 was discarded due to RMM misfit and VSR item 13 was discarded due to DIF based on gender. The remaining 14 VSR scale items were retained. The final measure order of these 14 VSR scale items (listed from easiest to most difficult) was as follows: 2, 3, 1, 8, 4, 9, 5, 6, 7, 15, 11, 12, 14 and 16 (see Fig. 3). The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the 14-item version of the TVPS-R VSR scale.

TVPS-R Visual Form Constancy Scale results

The TVPS-R Visual Form Constancy (VFC) scale RMM calibration findings are reported in Table 5. VFC scale item 1 was found to have RMM misfit. When the VFC scale items were examined for DIF based on gender, none of the items exhibited DIF.

The VFC scale items were found to have MNSQ infit statistics ranging from 0.86 to 1.18 and outfit statistics ranging from 0.47 to 1.22 (see Table 5). The mean VFC item measure was 0.00 logits (SD = 1.21). There was a moderately broad range of logit measures, the lowest value being -3.02 and the highest value being +1.87. This resulted in an item separation index of 7.15 and a reliability of 0.98. The TVPS-R VFC scale person-item map is located in Fig. 4. The average person measure was 1.21 logits (SD = 1.19). Person separation was 1.38, with a reliability of 0.66.

In summary, TVPS-R VFC scale item 1 was discarded because of RMM misfit, but the remaining 15 VFC scale items were retained. The final measure order of these 15 VFC scale items (listed from easiest to most difficult) was as follows: 2, 3, 5, 4, 11, 10, 14, 9, 15, 8, 12, 7, 6, 16 and 13 (see Fig. 4). The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the 15-item version of the TVPS-R VFC scale.

TVPS-R Visual Sequential Memory Scale results

The TVPS-R Visual Sequential Memory (VSM) scale items were calibrated using the RMM (see Table 6). The results indicated that the TVPS-R VSM scale items had infit and outfit statistics within the acceptable parameters. When the VSM scale items were then examined for DIF based on gender, item 3 exhibited DIF.

The TVPS-R VSM scale items were calibrated and were found to have MNSQ infit statistics ranging from 0.80 to 1.23 and outfit statistics ranging from 0.72 to 1.28 (see Table 6). The mean VSM item measure was 0.00 logits (SD = 1.03). There was a very limited range of logit measures. The lowest value was -1.68 and the highest value was only +1.51. This resulted in an item separation index of 6.82 and a reliability of 0.98. The TVPS-R VSM scale person-item map is located in Fig. 5. The average person measure was 1.12 logits (SD = 1.30). Person separation was 1.50, with a reliability of 0.

Once again, TVPS-R VSM scale item 3 exhibited DIF based on gender. The final measure order of the 15 VSM scale items (listed from easiest to most difficult) was as follows: 1, 5, 2, 4, 11, 9, 7, 12, 8, 6, 16, 15, 14, 10 and 13 (see Fig. 5). The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the 15-item version of the TVPS-R VSM scale.

TVPS-R Visual Figure Ground Scale results

The TVPS-R Visual Figure Ground (VFG) scale RMM calibration findings are reported in Table 7. The TVPS-R VFG calibration indicated that all of the scale items were found to have infit and outfit statistics within acceptable RMM parameters. Item 16 was found to exhibit DIF based on gender.

The VFG scale items were found to have MNSQ infit statistics ranging from 0.80 to 1.26 and outfit statistics ranging from 0.43 to 1.49 (see Table 7). VFG scale items 1, 3, 8 and 14 had low outfit scores whereas items 11 and 15 had high outfit scores. All of these items were retained since their infit scores fell within the 0.80 to 1.20 range. The mean VFG item measure was 0.00 logits (SD = 1.66). There was a broad range with respect to logit measures, the lowest value being -2.81 and the highest value being +3.67. This resulted in an item separation index of 6.89 and a reliability of 0.98. The TVPS-R VFG scale person-item map is located in Fig. 6. The average person measure was 1.92 logits (SD = 1.27). Person separation was 1.13, with a reliability of 0.56.

In summary, TVPS-R VFG scale item 16 exhibited DIF based on gender. The measure order of the VFG items (listed from easiest to most difficult) was as follows: 1, 8, 4, 2, 3, 12, 14, 5, 7, 10, 13, 6, 11, 9 and 15 (see Fig. 6). The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the 15-item version of the TVPS-R VFG scale.

TVPS-R Visual Closure Scale results

The 16 TVPS-R Visual Closure (VC) scale items were calibrated using the RMM. All of the VC scale items were found to have infit and outfit statistics within the acceptable parameters specified by the RMM (see Table 8). The VC scale items were then examined for DIF based on gender and VC item 2 was found to exhibit DIF.

The items were found to have MNSQ infit statistics ranging from 0.80 to 1.24 and outfit statistics ranging from 0.34 to 1.54 (see Table 8). VC scale items 1 and 5 had low outfit scores while VC scale items 7 and 14 had high outfit scores. All of these items were retained since their infit scores were in the 0.80 to 1.20 range. The mean VC item measure was 0.00 logits (SD = 1.45). There was a broad range with respect to logit measures, the lowest value being -4.09 and the highest value being +2.21. This resulted in an item separation index of 7.58 and a reliability of 0.98. The VC scale person-item map is located in Fig. 7. The average person measure was 1.14 logits (SD = 1.51). Person separation was 1.65, with a reliability of 0.73.

Table 5. TVPS-R Visual Form Constancy Scale item measure order (n = 356)

TVFC STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVFC
13	122	324	2.17	.13	1.15	1.6	1.22	1.1	.15	TVFC13
16	149	324	1.73	.13	1.18	2.0	1.19	1.2	.18	TVFC16
6	156	324	1.61	.13	1.05	.6	1.06	.4	.25	TVFC6
7	205	324	.81	.13	.92	−.9	.88	−.8	.40	TVFC7
12	209	324	.75	.13	1.12	1.2	1.17	1.0	.22	TVFC12
8	211	324	.71	.13	1.17	1.7	1.17	1.0	.17	TVFC8
15	223	324	.50	.14	.92	−.8	.82	−1.1	.40	TVFC15
9	228	324	.40	.14	.95	−.4	1.00	0.0	.37	TVFC9
14	228	324	.40	.14	.94	−.6	.83	−1.0	.39	TVFC14
10	230	324	.36	.14	.92	−.7	.88	−.6	.40	TVFC10
11	232	324	.33	.14	.86	−1.4	.70	−1.8	.47	TVFC11
4	251	324	−.07	.15	.94	−.5	.76	−1.1	.38	TVFC4
5	287	324	−1.06	.19	.93	−.4	.55	−1.3	.39	TVFC5
3	296	324	−1.43	.21	.99	0.0	.73	−.6	.29	TVFC3
2	314	324	−2.69	.34	.98	−.1	.46	−.7	.25	TVFC2
1	322	324	−4.52	.74	**1.20**	**.2**	**4.14**	**.8**	**−.02**	**TVFC1**
MEAN	229.	324.	.00	.19	1.01	.1	1.10	−.2		
S.D.	55.	0.	1.65	.15	.11	1.0	.82	1.0		

TVFC = TVPS-R Visual Form Constancy Scale item; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.
RMM misfitting item: 1.

Table 6. TVPS-R Visual Sequential Memory Scale item measure order (n = 356)

TVSM STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVSM
13	148	339	1.51	.12	1.08	1.0	1.08	.6	.28	TVSM13
10	150	339	1.48	.12	1.03	.3	1.06	.4	.30	TVSM10
14	153	339	1.43	.12	.95	−.7	.93	−.6	.36	TVSM14
15	190	339	.86	.13	1.22	2.6	1.24	1.9	.17	TVSM15
16	197	339	.75	.13	.96	−.5	.90	−.9	.39	TVSM16
6	207	339	.59	.13	1.07	.7	1.09	.7	.30	TVSM6
8	230	339	.20	.13	1.07	.7	.97	−.2	.31	TVSM8
12	245	339	−.08	.14	.80	−2.0	.73	−1.8	.53	TVSM12
3	256	339	−.30	.14	1.13	1.1	1.21	1.1	.24	TVSM3
7	260	339	−.38	.15	.90	−.9	.75	−1.4	.45	TVSM7
9	272	339	−.65	.15	.93	−.6	.75	−1.2	.42	TVSM9
11	272	339	−.65	.15	.98	−.2	.93	−.3	.37	TVSM11
4	274	339	−.70	.16	.97	−.2	.91	−.4	.37	TVSM4
2	288	339	−1.07	.17	.93	−.4	.74	−1.0	.41	TVSM2
5	297	339	−1.35	.18	.98	−.1	1.25	.7	.31	TVSM5
1	305	339	−1.64	.20	1.04	.2	.77	−.7	.31	TVSM1
MEAN	234.	339.	0.00	.15	1.00	.1	.96	−.2		
S.D.	52.	0.	.98	.02	.10	1.0	.18	1.0		

TVSM = TVPS-R Visual Sequential Memory Scale item; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients

Fig. 4. TVPS-R Visual Form Constancy Scale Rasch analysis person-item map (n = 356).

```
   CHILDREN                MAP OF TVFC Scale Items
          <frequ>|<less>
 4     ######## Q+
                 |
                 |
     .#########  |
                 |Q
                 |
 3               +
  .############ S|
                 |
                 |
                 |  TVFC13
 2   ##########  +
                 |
  .############  |S TVFC16  TVFC6
                M|
                 |
      #########  |
 1               +
        #######  |  TVFC7
                 |  TVFC12  TVFC8
         ######  |  TVFC15
                S|  TVFC10  TVFC11  TVFC14  TVFC9
          .####  |
 0               +M TVFC4
          .####  |
                 |
            ###  |
                 |
                 |
-1            . Q+  TVFC5
                 |
              #  |
                 |  TVFC3
                 |S
                 |
-2            .  +
                 |
                 |
                 |
                 |  TVFC2
                 |
-3               +
                 |
                 |Q
                 |
                 |
                 |
-4               +
              .  |
                 |
                 |  TVFC1
                 |
                 |
-5               +
           <rare>|<more>
```

TVFC = TVPS-R Visual Form Constancy Scale item.

Fig. 5. TVPS-R Visual Sequential Memory Scale Rasch analysis person-item map (n = 356).

```
   CHILDREN                MAP OF TVSM Scale Items
          <frequ>|<less>
 4         .###  +
                 |
                Q|
                 |
                 |
                 |
      .#######   |
 3               +
                 |
                 |
                 |
                S|
    ###########  |
                 |
 2               +Q
                 |
     .########   |
                 |  TVSM13
                 |  TVSM10  TVSM14
   .##########   |
                M|
 1      .######  +S
                 |  TVSM15
                 |  TVSM16
          .####  |  TVSM6
                 |
          .####  |
                 |  TVSM8
 0        .####  +M
                S|  TVSM12
          .####  |  TVSM3
                 |  TVSM7
            .##  |
                 |  TVSM11  TVSM4   TVSM9
                 |
-1             # +S TVSM2
                 |
              .  |  TVSM5
                Q|
                 |  TVSM1
              #  |
                 |
-2               +Q
                 |
              .  |
                 |
                 |
                 |
                 |
-3               +
              .  |
                 |
                 |
                 |
                 |
                 |
-4            .  +
           <rare>|<more>
```

TVSM = TVPS-R Visual Sequential Memory Scale item.

Fig. 6. TVPS-R Visual Figure Ground Scale Rasch analysis person-item map (n = 356).

```
   CHILDREN                MAP OF TFIG Scale Items
          <frequ>|<less>
 4 .############ +
                 |
        #######  |
                 |  TFIG15
                 |
                 |Q
                S|
 3               +
                 |
      .########  |
                 |
                 |
                 |
      .########  |
 2               +
                M|
                 |S TFIG9
        .######  |  TFIG11  TFIG16
                 |
                 |
          .####  |
 1               +
                 |  TFIG6
          .####  |
                S|  TFIG13
            .##  |
                 |  TFIG10  TFIG7
                 |
 0           ##  +M
                 |  TFIG12  TFIG14  TFIG5
                 |
             .#  |
                 |
              . Q|
                 |
-1               +  TFIG3
              .  |
                 |
                 |
                 |  TFIG2
                 |S TFIG4
                 |
-2               +
              .  |
                 |
                 |
                 |
                 |  TFIG8
                 |  TFIG1
-3               +
                 |
                 |Q
                 |
                 |
              .  |
                 |
-4               +
           <rare>|<more>
```

TFIG = TVPS-R Visual Figure Ground Scale item.

Table 7. TVPS-R Visual Figure Ground Scale item measure order (n = 356)

TFIG STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TFIG
15	59	281	3.56	.16	1.02	.2	1.46	1.1	.12	TFIG15
9	151	281	1.67	.14	1.00	-.1	.97	-.2	.29	TFIG9
16	153	281	1.63	.14	.88	-1.5	.84	-1.1	.39	TFIG16
11	158	281	1.54	.14	1.29	3.2	1.41	2.3	.08	TFIG11
6	191	281	.88	.15	1.03	.3	1.07	.4	.29	TFIG6
13	207	281	.52	.15	1.07	.7	1.34	1.5	.24	TFIG13
10	217	281	.28	.16	.98	-.2	.79	-1.0	.35	TFIG10
7	218	281	.25	.16	.95	-.4	.86	-.6	.36	TFIG7
5	230	281	-.08	.17	.94	-.4	.81	-.7	.37	TFIG5
14	232	281	-.14	.18	.81	-1.4	.59	-1.6	.49	TFIG14
12	234	281	-.21	.18	1.00	0.0	.98	-.1	.31	TFIG12
3	254	281	-.99	.22	1.10	.4	.74	-.6	.27	TFIG3
2	264	281	-1.60	.28	1.01	.1	.85	-.3	.30	TFIG2
4	265	281	-1.67	.28	1.00	.0	.83	-.3	.28	TFIG4
8	274	281	-2.73	.43	.80	-.4	.47	-.6	.34	TFIG8
1	275	281	-2.92	.46	.88	-.2	.52	-.5	.27	TFIG1
MEAN	211.	281.	0.00	.21	.98	0.0	.91	-.1		
S.D.	56.	0.	1.66	.10	.11	1.0	.28	1.0		

TFIG = TVPS-R Visual Figure Ground Scale item; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

Table 8. TVPS-R Visual Closure Scale item measure order (n = 356)

TVC STATISTICS: MEASURE ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZEMP	OUTFIT MNSQ	OUTFIT ZEMP	PTBIS CORR.	TVC .
14	102	311	2.28	.14	1.22	1.5	1.40	1.5	.25	TVC14
12	118	311	1.96	.14	.79	-1.7	.72	-1.6	.52	TVC12
11	150	311	1.35	.14	1.12	.9	1.13	.8	.37	TVC11
10	165	311	1.07	.14	.97	-.3	.93	-.4	.46	TVC10
15	174	311	.90	.14	.81	-1.6	1.03	.2	.56	TVC15
16	191	311	.57	.14	1.10	.8	1.14	.8	.36	TVC16
7	205	311	.30	.14	1.22	1.6	1.33	1.6	.28	TVC7
8	218	311	.03	.15	.86	-1.0	.81	-1.0	.52	TVC8
9	226	311	-.15	.15	1.07	.5	.88	-.5	.37	TVC9
13	236	311	-.38	.15	.93	-.4	.86	-.6	.45	TVC13
4	237	311	-.40	.15	.88	-.8	.80	-.8	.49	TVC4
5	244	311	-.58	.16	.89	-.7	.68	-1.2	.48	TVC5
3	251	311	-.76	.17	.91	-.5	.69	-1.1	.46	TVC3
6	258	311	-.96	.17	1.23	1.1	1.53	1.3	.20	TVC6
2	267	311	-1.24	.18	1.01	.1	1.37	.8	.31	TVC2
1	306	311	-3.97	.47	.95	-.1	.31	-.6	.21	TVC1
MEAN	209.	311.	.00	.17	1.00	0.0	.98	-.1		
S.D.	54.	0.	1.43	.08	.14	1.0	.31	1.0		

TVC = TVPS-R Visual Closure Scale; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

Fig. 7. TVPS-R Visual Closure Scale Rasch analysis person-item map (n = 356).

```
    CHILDREN                MAP OF TVC Scale Items
               <frequ>|<less>
4          #########  +
                      |
                      |
                      |
        .###########  |
                      |
                      |
3                     +
                      |Q
                     S|
             .######  |
                      |
                      |  TVC14
                      |
2             .#####  +  TVC12
                      |
                      |
                      |
              .#####  |S
                     M|  TVC11
             #######  |
1                     +  TVC10
                      |  TVC15
               .####  |
                      |  TVC16
              .#####  |
                      |  TVC7
                .###  |
0                     +M TVC8
                      |  TVC9
                .### S|
                      |  TVC13  TVC4
                .###  |  TVC5
                      |  TVC3
                      |
-1               .##  +  TVC6
                      |
                      |  TVC2
                  .#  |S
                      |
                     Q|
                   .  |
-2                    +
                      |
                      |
                      |
                   .  |
                      |
                      |Q
-3                    +
                      |
                      |
                      |
                      |
                   .  |
                      |
-4                    +  TVC1
                <rare>|<more>
```

TVC = TVPS-R Visual Closure Scale item.

In summary, although TVPS-R VC scale item 2 exhibited DIF based on gender, the remaining 15 VC scale items were retained. The final measure order of the VC scale items (listed from easiest to most difficult) was as follows: 1, 6, 3, 5, 4, 13, 9, 8, 7, 16, 15, 10, 11, 12 and 14 (see Fig. 7). The construct validity, scalability, hierarchical ordering and lack of DIF requirements were met by the final version of the TVPS-R VC scale.

TVPS-R motor-free visual perceptual composite scale (seven subscales combined)

The TVPS-R motor-free visual perceptual composite scale consists of 112 dichotomously scored items that make up seven subscales (for example, 16 items × 7 scales = 112 items). In the TVPS-R manual, only the seven individual subscales are summed together to get seven separate subscale summary scores. The seven subscales delimit the types of visual perceptual skills included in the TVPS-R. According to the TVPS-R manual, the 112 items from the seven subscales are not summed together to get one overall motor-free visual perceptual score. Instead, the seven standard scores derived from the seven scales item totals are summed to calculate an overall motor-free visual perceptual quotient.

The TVPS-R motor-free visual perceptual composite scale was calibrated in order to obtain the best fit for all of the scale items to the RMM. The results of the RMM analysis output are reported here in summary form only (see Fig. 8 and Table 9). The following TVPS-R composite scale items were found to misfit the RMM requirements: 28, 40, 49, 56, 61, 64, 77, 79, 81, 108, 5, 42, 54, 63, 72, 86 and 91. The following TVPS-R composite scale items exhibited DIF based on gender: 1, 34, 50, 51, 97, 98, 33, 35, 45 and 96. In all, 17 items exhibited RMM misfit and 10 items exhibited DIF based on gender.

When the remaining 85 TVPS-R motor-free visual perceptual composite scale items were calibrated, the items were found to have MNSQ infit statistics ranging from 0.81 to 1.24 and outfit statistics ranging from 0.47 to 1.47 (see Table 9). Items 20, 22, 25, 32, 67, 68, 69, 70, 78 and 93 all had high outfit statistics whereas items 2, 3, 10, 15, 38, 53, 62, 82, 83, 84, 87, 90, 94, 100, 101, 104, 109 and 111 all had low outfit statistics. These items were all retained since their infit scores were in the conventional range of 0.80 to 1.20. The mean item measure was 0.00 logits (SD = 1.18). With respect to logit measures, there was a broad range, with the lowest value being -4.27 and the highest value being +3.23. This resulted in an item separation index of 6.33 and a reliability

Fig. 8. TVPS-R Seven scales combined Rasch analysis person-item map.

```
   CHILDREN              MAP OF TVPS Scale Items
          <frequ>|<less>
 6                +
                  |
                  |
               #  |
                  |
 5                +
              ##  |
                  |
                 Q|
             .##  |
 4           .##  +
              .#  |
             .##  |
            ####  |
            .### S|  TVM16
 3        .######  +
            ####  |
          ######  |Q TFIG15
          ######  |  TVCL14
      .#########  |  TVFC13  TVM15
 2        .#####  +  TVCL12  TVSM13  TVSM10  TVSM14
       .######## M|  TVD16  TVFC16
          .#####  |  TVFC6  TVCL11
         .######  |S TVSM15  TVCL10  TVM13
       #########  |  TVSM16  TVCL15  TVSM6
 1        .#####  +  TFIG9  TFIG16  TFIG11  TVM14  TVD14  TVD9  TVCL16  TVFC7
           .####  |  TVFC12  TVD13  TVFC8  TVSM8  TVD10
            .### S|  TSR16  TVFC14  TVFC9  TVM11  TVFC15  TVM12  TVCL7
             .##  |  TFIG6  TSR14  TSR12  TVFC11  TVCL8  TVM8  TVFC10  TVSM12
             .##  |  TSR11  TVSM7  TVD11  TVSM3  TVCL9
 0          ####  +M TFIG13 TS15 TSR7 TVCL5 TVD4 TVSM11 TVSM9TVM10TVD12 TVFC4 TVCL4
                  |  TVM6 TVCL13
             .##  |  TFIG10  TFIG7  TSR6  TVM5  TVM7  TVCL3  TVM2  TVD8  TVSM4
            ####  |  TFIG5  TSR5  TSR10  TVD15  TVSM2  TVCL6  TVD6
              .#  |  TFIG14  TFIG12  TSR9  TVM4  TVM3  TVCL2
               # Q|  TVD7  TVSM5
-1              .  +  TSR13  TVD5  TVD2  TVSM1  TVFC5  TVD3
                  |  TSR4  TVFC3
               .  |S TFIG3  TSR8
                  |  TVM9
                  |  TFIG2
-2              .  +  TFIG4
                  |
                  |  TSR1  TSR3  TVFC2
                  |Q
                  |  TFIG8  TSR2  TVD1
-3                +  TFIG1
                  |  TVCL1
                  |  TVM1
                  |
                  |
-4                +
                  |  TVFC1
                  |
                  |
                  |
-5                +
            <rare>|<more>
```

TVD = TVPS-R Visual Discrimination Scale item; TVM = TVPS-R Visual Memory Scale item; TSR = TVPS-R Visual Spatial Relationships Scale item; TVFC = TVPS-R Visual Form Constancy Scale item; TVSM = TVPS-R Visual Sequential Memory Scale item; TFIG = TVPS-R Visual Figure Ground Scale item; TVCL = TVPS-R Visual Closure Scale item.

of 0.98. The TVPS-R motor-free visual perceptual scale person-item map is located in Fig. 8. The average person measure was 1.77 logits (SD = 1.41). Person separation was 3.39 with a reliability of 0.92.

In all, 27 out of 112, or 24%, of the TVPS-R motor-free visual perceptual scale items were found to either misfit or exhibit DIF. Seventeen items were discarded because of poor fit with the RMM (items 28, 40, 49, 56, 61, 64, 77, 79, 81, 108, 5, 42, 54, 63, 72, 86 and 91), and 10 items were discarded due to DIF (items 1, 34, 50, 51, 97, 98, 33, 35, 45 and 96) based on gender. In percentage terms, 15% of the items were excluded because of RMM misfit, and 9% of the items were excluded because of DIF based on gender.

It would appear that, since the items did not load on one dimension, the overall TVPS-R motor-free visual perceptual composite scale is a multidimensional construct. Also, given the fact that nearly 25% of the items were discarded due to RMM misfit or DIF, the current scale is not viable for use by practitioners. Therefore, the suggested practice of using scale quotients to calculate an overall visual perceptual scale score is not recommended. It would be better for professionals to use the seven scales and attempt to develop an overall profile of a client's visual perceptual abilities instead of relying on one overall summary score.

Discussion

In this study, all seven of the TVPS-R scales met the RMM requirements in terms of MNSQ infit and outfit statistics within the conventional limits for the t-statistic, in this case ranging from -0.80 to +1.20. Three TVPS-R scales had an item that exhibited RMM misfit (visual memory scale, visual spatial relationships scale and visual form constancy scale), and five scales had an item that exhibited DIF based on gender (visual memory scale, visual spatial relationships scale, visual sequential memory scale, visual figure ground scale and visual closure scale). Only the TVPS-R visual discrimination scale did not have any items that demonstrated either RMM misfit or DIF based on gender.

Klein et al (2002) recently completed both an exploratory factor analysis and a confirmatory factor analysis of the 1982 version of the TVPS-R with a sample of 294 children from Alberta, Canada. The seven scales were analysed separately so that the factor loadings for each scale could be examined. The results from the principal component analysis indicated that many, although not all, of the TVPS items across the seven subscales loaded on a dominant first factor. All of the items from the visual spatial relationships subscale consistently loaded on the first factor. Fifteen of the 16 items loaded on the first factor for the visual discrimination, visual sequential memory, visual figure ground and visual closure subscales, but only 12 of the 16 items loaded on the first factor from the visual form constancy and visual memory subscales. According to Klein et al (2002), items for some subscales consistently loaded on the first factor (for example, visual spatial relationships subscale), while other items from the other TVPS subscales appeared to be multidimensional since they loaded on across more multiple factors. This implies that there was not unidimensionality, but multiple factors in the subscales (for example, visual form constancy and visual memory subscales). These outcomes presented a complex structure where TVPS items loaded on multiple factors. As a consequence, Klein et al (2002) wondered whether the same factor was being measured by the different subscales.

The validity evidence presented in the TVPS-R manual is weak. Content validity was described in part as determining the significant factors of visual perception, but there was no discussion of the importance of selecting these factors. The evidence presented in the test manual

Table 9. TVPS-R Seven scales combined item measure order (n = 356)

TVPS STATISTICS: MEASURE ORDER

ENTRY	RAW				INFIT		OUTFIT		PTBIS	
NUMBER	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZEMP	MNSQ	ZEMP	CORR.	TVPS
32	92	356	3.22	.14	1.18	1.4	1.31	1.2	.27	TVM16
95	134	356	2.50	.13	.85	-1.6	.98	-.1	.47	TFIG15
110	147	356	2.30	.12	1.08	.8	1.19	1.2	.38	TVCL14
61	154	356	2.19	.12	**1.20**	**2.1**	**1.33**	**2.1**	**.30**	**TVFC13**
31	159	356	2.12	.12	.91	-1.0	.84	-1.2	.52	TVM15
108	163	356	2.06	.12	**.75**	**-3.0**	**.70**	**-2.4**	**.61**	**TVCL12**
77	164	356	2.04	.12	**1.25**	**2.5**	**1.30**	**2.0**	**.32**	**TVSM13**
74	166	356	2.01	.12	1.07	.8	1.13	.9	.40	TVSM10
78	169	356	1.97	.12	1.05	.6	1.18	1.2	.40	TVSM14
16	181	356	1.78	.12	1.07	.7	1.02	.1	.42	TVD16
64	181	356	1.78	.12	**1.30**	**3.0**	**1.49**	**3.1**	**.25**	**TVFC16**
54	188	356	1.68	.12	1.18	1.9	1.26	1.7	.33	TVFC6
107	195	356	1.57	.12	1.00	0.0	.98	-.1	.48	TVCL11
79	206	356	1.40	.12	**1.36**	**3.4**	**1.63**	**3.5**	**.21**	**TVSM15**
106	210	356	1.34	.12	.88	-1.3	.86	-.9	.54	TVCL10
29	212	356	1.31	.12	1.05	.5	1.21	1.3	.44	TVM13
80	213	356	1.30	.12	1.03	.3	1.01	.1	.45	TVSM16
111	219	356	1.20	.13	.81	-2.1	.72	-1.9	.62	TVCL15
70	223	356	1.14	.13	1.14	1.3	1.27	1.5	.36	TVSM6
89	226	356	1.09	.13	1.01	.1	.92	-.5	.46	TFIG9
30	228	356	1.06	.13	1.09	.9	1.11	.6	.40	TVM14
96	228	356	1.06	.13	.81	-2.0	.69	-2.0	.61	TFIG16
14	231	356	1.01	.13	.99	-.1	.91	-.5	.49	TVD14
9	232	356	.99	.13	1.02	.2	.96	-.3	.46	TVD9
91	233	356	.98	.13	1.22	2.0	1.18	1.0	.30	TFIG11
112	236	356	.93	.13	.98	-.2	.95	-.3	.48	TVCL16
55	237	356	.91	.13	1.08	.7	1.02	.1	.41	TVFC7
60	241	356	.84	.13	1.12	1.1	1.05	.2	.37	TVFC12
13	242	356	.83	.13	1.03	.3	.99	-.1	.45	TVD13
56	243	356	.81	.13	**1.31**	**2.6**	**1.32**	**1.5**	**.22**	**TVFC8**
72	246	356	.76	.13	1.19	1.6	1.34	1.5	.32	TVSM8
10	247	356	.74	.13	.86	-1.3	.75	-1.3	.58	TVD10
103	250	356	.69	.13	1.14	1.2	1.06	.3	.34	TVCL7
28	251	356	.67	.13	**1.27**	**2.2**	**1.52**	**2.1**	**.25**	**TVM12**
48	253	356	.64	.13	.99	-.1	1.13	.6	.47	TSR16
63	255	356	.60	.13	.98	-.2	.82	-.9	.48	TVFC15
27	257	356	.56	.13	.92	-.7	.99	0.0	.52	TVM11
57	260	356	.51	.14	.99	-.1	1.04	.2	.47	TVFC9
62	260	356	.51	.14	.96	-.3	.77	-1.1	.49	TVFC14
46	261	356	.49	.14	.96	-.4	.83	-.8	.50	TSR14
76	261	356	.49	.14	.90	-.9	.85	-.7	.54	TVSM12
58	262	356	.47	.14	1.04	.4	1.04	.2	.42	TVFC10
24	263	356	.45	.14	.92	-.6	.96	-.2	.51	TVM8
104	263	356	.45	.14	.81	-1.7	.67	-1.6	.61	TVCL8
59	264	356	.43	.14	.90	-.8	.77	-1.0	.53	TVFC11
86	266	356	.40	.14	1.06	.5	1.41	1.5	.40	TFIG6
44	269	356	.34	.14	.90	-.8	.80	-.8	.53	TSR12
105	271	356	.30	.14	1.04	.3	.87	-.5	.42	TVCL9
67	272	356	.28	.14	1.20	1.4	1.28	1.0	.28	TVSM3
11	274	356	.24	.14	.93	-.6	.88	-.5	.50	TVD11
71	276	356	.20	.14	.99	-.1	.90	-.4	.45	TVSM7
43	279	356	.14	.14	1.00	0.0	.88	-.5	.44	TSR11
109	281	356	.09	.15	.93	-.5	.71	-1.1	.50	TVCL13
22	282	356	.07	.15	1.20	1.3	1.28	.9	.27	TVM6
47	282	356	.07	.15	.91	-.6	.87	-.5	.50	TSR15
93	282	356	.07	.15	1.08	.5	1.18	.6	.37	TFIG13
100	282	356	.07	.15	.89	-.8	.70	-1.1	.53	TVCL4
52	283	356	.05	.15	.95	-.4	.83	-.6	.48	TVFC4
12	284	356	.03	.15	1.11	.8	.96	-.1	.35	TVD12
39	284	356	.03	.15	.90	-.7	.78	-.8	.52	TSR7
26	285	356	.01	.15	.93	-.5	.94	-.2	.49	TVM10
73	288	356	-.06	.15	.93	-.5	.78	-.8	.49	TVSM9
75	288	356	-.06	.15	.99	-.1	.84	-.5	.44	TVSM11
4	289	356	-.08	.15	.96	-.3	.79	-.7	.47	TVD4
101	289	356	-.08	.15	.92	-.5	.71	-1.0	.50	TVCL5
68	290	356	-.11	.15	1.03	.2	1.06	.2	.40	TVSM4
38	292	356	-.15	.15	.91	-.6	.73	-.9	.50	TSR6
90	292	356	-.15	.15	.96	-.2	.75	-.8	.46	TFIG10
21	293	356	-.18	.15	1.03	.2	1.20	.6	.38	TVM5
87	293	356	-.18	.15	.91	-.6	.74	-.9	.50	TFIG7
23	296	356	-.25	.16	.99	-.1	.82	-.5	.43	TVM7
99	296	356	-.25	.16	.92	-.5	.77	-.7	.49	TVCL3
18	297	356	-.27	.16	.99	0.0	.90	-.3	.41	TVM2
8	298	356	-.30	.16	.87	-.8	.71	-.9	.52	TVD8
6	300	356	-.35	.16	1.03	.2	.91	-.2	.38	TVD6
37	303	356	-.43	.16	.92	-.5	.88	-.3	.47	TSR5
42	303	356	-.43	.16	.78	-1.4	.49	-1.6	.59	TSR10
102	303	356	-.43	.16	1.17	.9	1.07	.2	.26	TVCL6
66	304	356	-.46	.17	.92	-.5	.77	-.6	.46	TVSM2
15	305	356	-.48	.17	.81	-1.1	.68	-.9	.55	TVD15
85	305	356	-.48	.17	.96	-.2	.85	-.4	.43	TFIG5
94	307	356	-.54	.17	.83	-1.0	.59	-1.2	.54	TFIG14
41	308	356	-.57	.17	.96	-.2	.87	-.3	.42	TSR9
92	309	356	-.60	.17	.97	-.1	.78	-.6	.41	TFIG12
20	311	356	-.66	.17	1.14	.7	1.27	.6	.25	TVM4
19	312	356	-.69	.18	.97	-.2	1.16	.3	.38	TVM3
98	312	356	-.69	.18	1.04	.2	1.01	0.0	.35	TVCL2
7	313	356	-.72	.18	1.14	.7	.96	-.1	.26	TVD7
69	313	356	-.72	.18	1.07	.3	1.44	.9	.30	TVSM5
3	319	356	-.92	.19	.82	-.9	.50	-1.2	.52	TVD3
53	319	356	-.92	.19	.87	-.6	.54	-1.1	.49	TVFC5
65	321	356	-.99	.19	.98	-.1	.92	-.2	.37	TVSM1
2	322	356	-1.03	.19	.96	-.2	.69	-.7	.39	TVD2
5	322	356	-1.03	.19	.89	-.5	.78	-.4	.46	TVD5
45	323	356	-1.07	.20	.84	-.7	.52	-1.1	.49	TSR13
36	328	356	-1.27	.21	.97	-.1	.82	-.3	.35	TSR4
51	328	356	-1.27	.21	1.00	.0	.61	-.8	.34	TVFC3
83	329	356	-1.32	.21	1.05	.2	.71	-.5	.30	TFIG3
40	330	356	-1.36	.22	**.87**	**-.5**	**.40**	**-1.2**	**.47**	**TSR8**
25	334	356	-1.56	.23	1.05	.2	1.44	.6	.21	TVM9
82	339	356	-1.86	.26	.89	-.3	.61	-.6	.37	TFIG2
84	340	356	-1.93	.27	.94	-.2	.66	-.5	.30	TFIG4
33	345	356	-2.35	.32	.91	-.2	.59	-.5	.32	TSR1
35	346	356	-2.46	.33	.99	0.0	.66	-.3	.26	TSR3
50	346	356	-2.46	.33	.95	-.1	.68	-.3	.24	TVFC2
1	348	356	-2.70	.37	.98	0.0	.85	-.1	.24	TVD1
34	349	356	-2.84	.39	.98	0.0	.46	-.5	.24	TSR2
88	349	356	-2.84	.39	.91	-.1	.88	-.1	.24	TFIG8
81	350	356	-3.01	.42	**.90**	**-.2**	**.34**	**-.6**	**.27**	**TFIG1**
97	351	356	-3.20	.46	.97	0.0	.23	-.7	.25	TVCL1
17	352	356	-3.43	.51	1.02	0.0	1.28	.1	.11	TVM1
49	354	356	-4.15	.71	**1.03**	**0.0**	**9.90**	**1.6**	**-.02**	**TVFC1**
MEAN	272.	356.	0.00	.18	1.00	0.0	1.00	-.1		
S.D.	55.	0.	1.34	.09	.12	1.0	.89	1.0		

TVD = TVPS-R Visual Discrimination Scale item; TVM = TVPS-R Visual Memory Scale item; TSR = TVPS-R Visual Spatial Relationships Scale item; TVFC = TVPS-R Visual Form Constancy Scale item; TVSM = TVPS-R Visual Sequential Memory Scale item; TFIG = TVPS-R Visual Figure Ground Scale item; TVCL = TVPS-R Visual Closure Scale item; Measure = item logit score; MNSQ = mean square; ZEMP = z-score; PTBISCORR. = point biserial correlation coefficients.

RMM misfitting items: TVM12 (item 28); TSR8 (item 40); TVFC1 (item 49); TVFC8 (item 56); TVFC13 (item 61); TVFC16 (item 64); TVSM13 (item 77); TVSM15 (item 79); TFIG1 (item 81); and TVCL12 (item 108).

for diagnostic validity is the low subscale intercorrelations and the differences in TVPS-R subscores between a subject group with average capabilities and a subject group with known learning disabilities. Although significant, there was no basis for predicting on which subscales specific types of learning disabled participants will differ from the normative sample group. A similar situation exists for predictive validity, where there is no discussion as to why visual perception should predict achievement scores and outcomes (Burtner et al 1997).

The evidence for criterion validity reported in the TVPS-R manual is of poor quality as well, since it suffers from the debate over just what the appropriate criterion for a test of visual perception should be. Gardner (1996) used chronological age, the picture completion test of the *Wechsler Intelligence Scale for Children – third edition*, the *Bender Visual-Motor Gestalt Test* and the *Developmental Test of Visual-Motor Integration*. However, Busch-Rossnagel (1985) believed that the use of these criteria exhibited circuitous reasoning. The items on the TVPS-R were selected to correlate with chronological age, so the relatively high correlations are not evidence of criterion validity. Gardner (1982) viewed the relatively low correlations with the other instruments as evidence of the unique variance of the TVPS. Even though this may be considered evidence of content validity, it demonstrates the lack of a realistic criterion, which, according to Busch-Rossnagel (1985), is problematic for all such instruments. For the TVPS, the lack of a criterion is also related to the lack of a clear purpose: 'Given that the hidden assumption behind its development appears to be the diagnosis of learning difficulties, the author may be well advised to concentrate on evidence of diagnostic and predictive validity' (Busch-Rossnagel 1985, p1595).

Denison (1985), in fact, believed that validity considerations emerged as the weakest aspect of the TVPS. For example, no research-based rationale was developed for the particular construct of visual perception or the subcomponents that were outlined as part of the TVPS. The decision not to discuss the choice of two-dimensional shapes and figures over other potential stimuli was, as Denison (1985, p1597) put it, '[a] noteworthy omission'. Nor was a rationale provided for the role of visual skills (for example, visual-motor skills, eye-hand coordination, visual tracking skills and manual dexterity) in the TVPS manual.

In addition, since no factor analysis procedures were included in the manual, there is no way to assess empirically the discreteness or commonality of the seven subscales. Yet from an inspection of the subscales themselves, it would appear that both the form constancy and figure ground subscales contain items that are embedded or hidden by distracting features. Success on the visual sequential memory subscale items would also appear to depend on abilities in visual spatial type skills. According to Denison (1985), '[t]hese concerns deserve a much more complete consideration, especially since the efforts to fractionate the interactions between the environment and the person, with the corresponding reification of skills through a labeling of the various types of stimuli, have been questioned on both theoretical and practical bases' (p1597).

Denison (1985) suggested that the main problem that the TVPS manual does not address is not so much one of whether there are relationships between various visual perceptual skills, or between these skills and learning, as it is one of how well the TVPS has clarified or actually tapped these relationships. The information available in the TVPS-R manual provides very little of a definitive or inclusive nature in this respect. Consequently, too much is left up to the discretion of the test user when deciding what the interpretation of a participant's performance really means.

In summary, the TVPS-R offers a set of scales that claims to measure motor-free visual perception, a complicated, multidimensional construct. Yet the various scores promote a sense of utility and reality regarding the TVPS-R, which has not been established in the literature or sufficiently addressed by the test author: 'Although some basic and beginning steps have been taken to study the diagnostic and predictive implications of the TVPS, no conclusions can be drawn about what is gained by its administration' (Denison 1985, p1597). Gardner (1996) did not address the theoretical construct of visual perception nor did he justify the inclusion of the distinct visual perceptual abilities that are purported to represent this construct. The seven subscales or factors were developed, but no factor analysis was presented that would support empirically the existence of seven discrete visual perceptual abilities. The absence of a more thorough consideration of the construct upon which the TVPS-R is based, upon the meaning of various performance levels and upon its educational ramifications are all major limitations of this instrument. The findings in this study contribute to this body of knowledge.

When the final revised versions of the seven subscales were considered on an individual basis in this study, they all met the RMM construct validity, unidimensionality, scalability, hierarchical ordering and lack of DIF requirements. However, the measure order of the seven scales determined by the RMM were all markedly different from those presented in the TVPS-R test plate books. Again, the performance of participants completing the seven TVPS-R scales may be compromised if, as suggested by these results, they are required to complete more difficult items before easier items. Thus, even though the seven visual perceptual scales for the most part did meet the measurement requirements of the RMM and should be considered viable individual subscales, the measure order of the items needs to be examined more closely during their next revision. One factor to consider is that the results of this investigation are only generalisable to groups of children with typical development between 5 and 12 years of age.

The seven scales were also combined into one overall visual perceptual composite scale comprising the 112 items. Seventeen items were found to have poor fit within the conventional limits of -0.80 to +1.20 for the RMM MNSQ infit and outfit t-statistics. Several items had outfit scores that fell outside the 0.80 to 1.20 range, but were retained since their infit scores fell within the conventional limits. Ten items exhibited DIF based on gender. A final revised measure order of the remaining 85 TVPS-R motor-free visual perceptual scale items was proposed based on these results.

The final version of the 85 items fits together to measure an overall construct of motor-free visual perception. The one disadvantage of merging the items from seven different scales together into one overall motor-free visual perceptual scale is that it forces the assessor constantly to change the instructions given to the participant for completing each item (for example, completing a visual memory scale item, then completing a visual figure ground scale item). Practically speaking, this would be onerous for the test administrator and confusing for the respondent completing the scale.

Given the fact that the final version of the seven scales individually meet the RMM and DIF requirements, it makes more sense for the participants to complete the seven scales on an individual basis rather than have the examiner administer all of the items together as one overall motor-free visual perceptual scale. The use of separate subscales, each of which has been shown to be unidimensional, is further confirmed by the suggestion that the motor-free visual perceptual construct, as measured by the TVPS-R, is multidimensional rather than unidimensional, as required by the RMM.

Klein et al (2002), using a confirmatory factor analysis, evaluated the factor structure for all seven subscales at the same time by fitting the subscale scores to a unidimensional model. Four types of fit indices were used: the chi-square statistic, the root-mean square error of approximation, the root-mean square residual and the adjusted goodness-of-fit index. The results from the confirmatory analysis did not support the unidimensional assumption because the one-factor model provided poor fit to the observed data. In other words, the goodness-of-fit indices were beyond the acceptable level of fit, indicating that the TVPS scale data did not fit the hypothesised model (Klein et al 2002). According to Klein et al (2002), the confirmatory factor analysis did not support the use of the TVPS perceptual quotient as representing a unidimensional measure of visual perception. Because of this, they cautioned clinicians in their use and interpretation of the TVPS perceptual quotient.

Limitations and future studies

This study had several limitations. First, only the motor-free visual perception theoretical construct was considered. Secondly, the TVPS-R was developed by a researcher in the United States. As a result, the normative data for the TVPS-R were based only on American participants. Thirdly, only children presenting with normal profiles were included as participants in this study. Children with either an intellectual or a physical impairment (including developmental delays and/or learning disabilities) or not possessing a working knowledge of the English language were excluded. Finally, when recruiting children as participants, only those who, along with their parents, consented to participate in the study were included. There is always the possibility of some bias related to parental consent. Children whose parents did not provide consent may possess some unique characteristics or score profiles.

It would be worthwhile to evaluate the discriminant/diagnostic validity of the TVPS-R by determining its ability to differentiate between a group of respondents who have a clinical diagnosis (such as cerebral palsy or attention deficit disorder) and a group who are developing typically. Another suggestion for future study is to evaluate the predictive validity of the TVPS-R. For example, is the TVPS-R able to predict the future academic abilities of school-age children?

Conclusion

When the seven individual TVPS-R scales were analysed using the RMM, they all exhibited adequate measurement properties (for example, scalability/interval level measurement, unidimensionality, lack of DIF, and hierarchical ordering of items). However, when they were collapsed together to form an overall composite scale of motor-free visual perceptual skills, the items failed to group together to measure a unidimensional construct. In addition, many scale items exhibited RMM misfit or DIF. These results suggest that the seven TVPS-R scales can be used on an individual basis with clients to generate a profile of their motor-free visual perceptual skills, but that they cannot be compiled together to calculate an overall summary motor-free visual perceptual score or perceptual quotient. In summary, the construct validity of the seven individual TVPS-R scales was supported, but the construct validity of the composite TVPS-R instrument was not supported.

Acknowledgements

Acknowledgements are extended to three organisations whose funding grants made this study possible: (1) the Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; (2) the Bloorview Children's Hospital Foundation, Toronto, Ontario, Canada; and (3) the Canadian Occupational Therapy Foundation, Ottawa, Ontario, Canada. Thanks are also extended to the children who volunteered to take part in the study by completing the visual perceptual instruments.

Key findings

- The seven TVPS-R subscales can be used on an individual basis to generate a profile of visual perceptual skills.
- The TVPS-R composite scale does not exhibit adequate construct validity.

What the study has added

The construct validity of the seven individual TVPS-R scales was supported; however, the construct validity of the composite TVPS-R instrument was not.

References

American Occupational Therapy Association (1991) Statement: Occupational therapy provision for children with learning disabilities and/or mild to moderate perceptual and motor deficits. *American Journal of Occupational Therapy, 45,* 1069-73.

American Occupational Therapy Association (1994) Uniform terminology for occupational therapy. *American Journal of Occupational Therapy, 48,* 1047-54.

American Occupational Therapy Association (1997) *Occupational therapy services for children and youth under the Individuals with Disabilities Education Act.* Bethesda, MD: AOTA.

American Occupational Therapy Association (2002) Occupational therapy practice framework: domain and process. *American Journal of Occupational Therapy, 56,* 609-39.

Bishop K, Curtin M (2001) The TVPS, MVPT and VMI: what influences a therapist's choice? *National Association of Paediatric Occupational Therapists Journal, 5,* 8-11.

Bond TG, Fox CM (2001) *Applying the Rasch Model: fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bond TG, Fox CM (2007) *Applying the Rasch Model: fundamental measurement in the human sciences (second edition).* Mahwah, NJ: Lawrence Erlbaum Associates.

Brown GT, Gaboury I (2006) The measurement properties and factor structure of the Test of Visual Perceptual Skills – Revised: implications for occupational therapy assessment and practice. *American Journal of Occupational Therapy, 60,* 182-93.

Brown GT, Rodger S, Davis A (2003) Test of Visual Perceptual Skills – Revised: a review and critique. *Scandinavian Journal of Occupational Therapy, 10,* 3-15.

Brown GT, Rodger S, Brown A, Roever C, (2005) A comparison of Canadian and Australian paediatric occupational therapy practice: theory, interventions, and assessments. *Occupational Therapy International, 12,* 137-61.

Burtner P, Whilhite C, Bordegaray J, Moedl D, Roe R, Savage A (1997) Critical review of visual perceptual tests frequently administered by pediatric therapists. *Physical and Occupational Therapy in Pediatrics, 17,* 39-61.

Busch-Rossnagel NA (1985) Review of the Test of Visual Perceptual Skills (Non-Motor). In: JV Mitchell, ed. *The ninth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurement, University of Nebraska – Lincoln, University of Nebraska Press, 1595-96.

Chaikin LE, Downing-Baum S (1997) Functional visual skills. In: M Gentile, ed. *Functional visual behavior: a therapist's guide to evaluation and treatment options.* Rockville, MD: American Occupational Therapy Association, 105-32.

Chia SH (1997) Occupational therapists' assessment practices with children who have disabilities. *British Journal of Therapy and Rehabilitation, 4,* 123-28.

Chu S, Chia SH (1997) A review of assessments used in paediatric occupational therapy. *British Journal of Therapy and Rehabilitation, 4,* 228-33.

Colarusso RP, Hammill DD (1996) *Motor-Free Visual Perception Test – Revised.* Novato, CA: Academic Therapy Publications.

Crowe TK (1989) Pediatric assessments: a survey of their use by occupational therapists in northwestern school systems. *Occupational Therapy Journal of Research, 9,* 273-86.

Daniels LE, Wong K (1993) Visual perceptual and visual motor performance differences in children with learning disabilities. *Journal of Special Education, 17,* 289-93.

Dankert HL, Davies PL, Gavin WJ (2003) Occupational therapy effects on visual-motor skills in preschool children. *American Journal of Occupational Therapy, 57,* 542-49.

Denison JW (1985) Review of the Test of Visual Perceptual Skills (Non-Motor). In: JV Mitchell, ed. *The ninth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurement, University of Nebraska – Lincoln, University of Nebraska Press, 1596-98.

Erhardt RP, Duckman RH (2005) Visual-perceptual-motor dysfunction and its effects on eye-hand coordination and skill development. In: M Gentile, ed. *Functional visual behaviour in children: an occupational therapy guide to evaluation and treatment options.* Rockville, MD: American Occupational Therapy Association, 171-228.

Feder KP, Majnemer A, Synnes A (2000) Handwriting: current trends in occupational therapy practice. *Canadian Journal of Occupational Therapy, 67,* 197-204.

Fischer GH, Molenaar IW (1995) *Rasch models: foundations, recent developments, and applications.* New York: Springer.

Fischer B, Hartnegg K, Mokler A (2000) Dynamic visual perception of dyslexic children. *Perception, 29,* 523-30.

Gardner MF (1982) *Test of Visual-Perceptual Skills (Non-Motor).* San Francisco, CA: Psychological and Educational Publications.

Gardner MF (1996) *Test of Visual Perceptual Skills (Non-Motor) – Revised.* San Francisco, CA: Psychological and Educational Publications.

Gentile M (1997) *Functional visual behavior: a therapist's guide to evaluation and treatment options.* Rockville, MD: American Occupational Therapy Association.

Grieve J (2000) *Neuropsychology for occupational therapists: assessment of perception and cognition.* Malden, MA: Blackwell Science.

Groffman S, Solan H (1994) *Developmental and perceptual assessment of learning-disabled children: theoretical concepts and diagnostic testing.* Santa Ana, CA: Optometric Extension Program Foundation.

Hammill DD, Pearson NA, Voress JK (1993) *Developmental Test of Visual Perception – second edition.* Austin, TX: ProEd.

Hart DL, Velozo CA, Lai JS, Dobrzykowski EA (1997) The reliability and validity of the Orthopedic Rehabilitation Outcome Scale. *Journal of Rehabilitation Outcomes Measurement, 1,* 1-7.

Hung SS, Fisher AG, Cermak SA (1987) The performance of learning-disabled and normal young men on the Test of Visual-Perceptual Skills. *American Journal of Occupational Therapy, 41,* 790-97.

Karabatos G (2001) The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement, 2,* 389-423.

Kirkpatrick LA, Feeney BC (2001) *A simple guide to SPSS for Windows for versions 8.0, 9.0 and 10.0.* Scarborough, ON: Nelson/Thomson Learning.

Klein S, Sollereder P, Gierl M (2002) Examining the factor structure and psychometric properties of the Test of Visual-Perceptual Skills. *Occupational Therapy Journal of Research: Occupation, Participation and Health, 22,* 16-24.

Kovacs I (2000) Human development of perceptual organization. *Vision Research, 40,* 1301-10.

Linacre JM, Wright BD (1998) *A user's guide to Bigsteps Winsteps Rasch-Model computer program.* Chicago, IL: MESA Press.

Loikith CC (2005) Development of visual attention. In: M Gentile, ed. *Functional visual behaviour in children: an occupational therapy guide to evaluation and treatment options.* Rockville, MD: American Occupational Therapy Association, 145-70.

Parush S, Yochman A, Cohen D, Gershon E (1998) Relation of visual perception and visual-motor integration for clumsy children. *Perceptual and Motor Skills, 86,* 291-95.

Richardson PK (2005) Use of standardized tests in pediatric practice. In: J Case-Smith, ed. *Occupational therapy for children.* St Louis, MO: Elsevier Mosby, 246-75.

Rodger S (1994) A survey of assessments used by paediatric occupational therapists. *Australian Occupational Therapy Journal, 41,* 137-42.

Scheiman M (1997) *Understanding and managing visual deficits: a guide for occupational therapists.* Thorofare, NJ: Charles B Slack.

Schneck CM (2001) Visual perception. In: J Case-Smith, ed. *Occupational therapy for children.* Toronto, ON: Mosby, 382-412.

Schneck CM, Lemer PS (1993) Reading and visual perception. In: CB Royeen, ed. *AOTA self study series: Classroom applications for school-based practice.* Rockville, MD: American Occupational Therapy Association, 1-48.

Smith RM (1991) The distribution properties of Rasch item fit statistics. *Educational and Psychological Measurement, 51,* 541-65.

Smith RM (1992) *Applications of Rasch measurement.* Chicago, IL: MESA Press.

Smith EV, Smith RM (2004) *Introduction to Rasch measurement.* Maple Grove, MN: JAM Press.

Solan HA, Ciner EB (1989) Visual perception and learning: issues and answers. *Journal of the American Optometric Association, 60,* 457-60.

Tennant A, Conaghan PG (2007) The Rasch Measurement Model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis and Rheumatism (Arthritis Care and Research), 57(8),* 1358-62.

Todd VR (1993) Visual perceptual frame of reference: an information processing approach. In: P Kramer, J Hinojosa, eds. *Frames of reference for pediatric occupational therapy.* Baltimore, MD: Williams and Wilkins, 177-232.

Van Waelvelde H, De Weerdt W, De Cock P, Smits-Engelsman BC (2004) Association between visual perceptual deficits and motor deficits in children with developmental coordination disorder. *Developmental Medicine and Child Neurology, 46,* 661-66.

Velozo CA, Kielhofner G, Lai J-S (1999) Quantitative Research Series – The use of Rasch analysis to produce scale-free measurement of functional ability. *American Journal of Occupational Therapy, 53,* 83-90.

Weil MJ, Amundson SJC (1994) Relationship between visuomotor and handwriting skills of children in kindergarten. *American Journal of Occupational Therapy, 48,* 982-88.

Werner P, Rini L (1976) *Perceptual-motor development equipment.* New York: John Wiley.

Wright BD, Masters GN (1982) *Rating scale analysis.* Chicago, IL: Mesa Press.

Wright BD, Stone MH (1979) *Best test design.* Chicago, IL: Mesa Press.

Wright BD, Linacre JM, Heinemann AW (1993) Measuring functional status in rehabilitation. *Physical Medicine and Rehabilitation Clinics of North America, 4,* 475-91.

Wright BA, Bowen RW, Zecker SG (2000) Nonlinguistic perceptual deficits associated with reading and language disorders. *Current Opinion in Neurobiology, 10,* 482-86.