

Reliability and validity of the Chinese version of the psycho-educational profile (third edition) performance test

Daniel TL Shek^{1-5,*}, PhD, FHKPS, BBS, JP, and Lu Yu¹, PhD

¹Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, PR China

²Centre for Innovative Programme for Adolescents and Families, The Hong Kong Polytechnic University, Hong Kong, PR China

³Department of Social Work, East China Normal University, Shanghai, PR China

⁴Kiang Wu Nursing College of Macau, Macau, PR China

⁵Division of Adolescent Medicine, Department of Pediatrics, Kentucky Children's Hospital, University of Kentucky College of Medicine, Lexington, Kentucky, USA

Abstract

The Psycho-Educational Profile-3rd edition (PEP-3) is a popular instrument that assesses the skills and behaviors of children with autism and communicative disabilities, while its application in Chinese population has not been systematically examined. Based on the data collected from a clinical sample of children ($N = 455$) diagnosed as having autism spectrum disorder (ASD) and other pervasive developmental disorders (PPDs) as well as a sample of healthy controls ($N = 281$) in Hong Kong, this study examined the psychometric properties of the Performance Test of the Chinese version of the PEP-3 (CPEP-3). Different types of evidence on the reliability (internal consistency, test-retest reliability and inter-rater reliability) and validity (content validity and criterion-related validity) of the test were examined. It was found that the Performance Test of CPEP-3 possessed sound psychometric properties in assessing the developmental level of Chinese children with autism spectrum disorder in Hong Kong. The translated CPEP-3 Performance Test is a reliable and valid measure for assessing Chinese children with autism.

Keywords: autistic spectrum disorder, Chinese, psycho-educational profile-3rd edition, psychometric properties

Introduction

Autism spectrum disorder (ASD) refers to a group of developmental disabilities, characterized by unreciprocated interpersonal and emotional interactions, disordered language and communication, and repetitive and stereotypic behaviors (1-3). Although earlier studies had suggested the prevalence of autism was low, the global prevalence of ASD increased tremendously in the global context in the past decades (4-8). While children with ASD share similar symptoms, the age of onset, nature and severity of the problems are highly variable, which require different

* **Correspondence:** Professor Daniel TL Shek, PhD, FHKPS, BBS, JP, Chair Professor of Applied Social Sciences, Faculty of Health and Social Sciences, Department of Applied Social Sciences, The Hong Kong Polytechnic University, Room HJ407, Core H, Hunghom, Hong Kong. E-mail: daniel.shek@polyu.edu.hk

treatment approaches (9). Hence, psychometrically sound instruments that can help professionals to objectively assess children with ASD are of great importance to ensure accurate diagnosis and appropriate educational programming (10).

According to Shriver, Allen and Mathew (11), assessment is considered a core component in treating children with ASD in terms of three aspects. First, assessment helps to obtain an accurate verification for the diagnosis and special education eligibility. Second, assessment can guide the development of appropriate interventions by providing a descriptive profile on children's development in different areas. Third, assessment enables researchers to establish baselines against which evaluation of different intervention programs could be conducted.

Among different types of assessments measuring children with ASD, the Psycho-Educational Profile-3rd edition (PEP-3) is a standardized observation tool primarily designed for planning educational program for children with ASD from 2 to 7 years of age (12). It is the revised version of the Psycho-Educational Profile-Revised (PEP-R) which has been widely used in different parts of the world for the assessment of children with ASD in terms of their developmental skills and behaviors (13-17). Compared with the PEP-R, there are several significant improvements in the PEP-3. First, the measured function domains have been revised by incorporating the most recent research findings, such as the social and communication aspects. Second, normative data from both a sample with autism and a typically developing sample were provided which makes it possible to compare the two groups of children on their developmental profiles. Third, a Caregiver Report, which utilizes information reported by parents and other caregivers about the child's development and behavior in daily lives, has been added to help professionals have a fuller understanding about the child. Fourth, new items and important subtests were added into the PEP-3 while obsolete items were deleted. The validity of the revised instrument has been examined in the West (18).

Specifically, the PEP-3 consists of two complementary parts. The first part is a norm-referenced performance scale (Performance Test) which measures autistic children's development in three areas: communication, motor skills, and

maladaptive behaviors. The second part is a parent or caregiver reported form (Caregiver Report). Children's development and performance in communication, motor and social functioning, adaptive behaviors and thinking are rated by their caregivers based on their daily observation. As autistic children's symptoms vary across settings (such as people, places, and things the child interacts with), using both professional observers' ratings on children's performance in the laboratory and caregivers' ratings on children's daily lives helps to obtain a richer and comprehensive developmental profile of the child with reference to the strengths and weaknesses of the child in different environments. This could be very useful for professionals to design individualized educational and behavioral intervention programs in different settings.

Based on a normative sample of 407 children with ASD and 148 children with typical development in the United States, Schopler et al. (18) reported the reliability and validity of the PEP-3 in its user manual. As far as the reliability is concerned, the PEP-3 showed high reliability, with all subtests showing high internal consistency (alphas ranged from 0.84 to 0.90) and temporal stability (test-retest coefficients over a period of two weeks ranged from 0.94 to 0.99). Besides, for the inter-rater reliability for the Caregiver Report, the polychoric correlation coefficients ranged from 0.52 to 1.00. In terms of the validity, consistent with the expectations, autistic children's scores on PEP-3 Performance subtests were negatively correlated with their scores on measures of symptom severity in autistic children, including the Childhood Autism Rating Scales (CARS) (19) and the Autism Behavior Checklist-Second Edition (ABC) (20). For the Caregiver Report, positive correlation coefficients were found between the Vineland Adaptive Behavior Scales (VABS) (21) which assess parent-reported adaptive behavior of children and PEP-3. These findings provide sound support for the good psychometric properties of the PEP-3 when applied to children with ASD in the United States.

The PEP-R has been translated into several languages and research findings supporting the psychometric properties of the translated measures have been reported (14, 15, 22). In contrast, there are very few reports on the application of the PEP-3 in countries other than the United States (23). As

Chinese people constitute roughly one-fifth of the world's population, translation of PEP-3 to Chinese and establishment of the related psychometric properties should be conducted. Actually, a Chinese version of the PEP-R (CPEP-R) was validated and used in assessing children with ASD in Hong Kong since 2005 which has largely facilitated the diagnoses and treatment of autistic children (14). Nevertheless, very few reports have been published on the Chinese version of PEP-3 and its psychometric properties. In Taiwan, researchers translated the PEP-3 into Chinese and reported the reliability and validity of the instrument based on Taiwanese children with autism. While the preliminary findings supported the application of PEP-3 in Chinese children, the published studies were mostly based on small samples (ranged from 42 to 64 participants), which makes the generalization of the findings to other Chinese population questionable (22-24).

Against the above background, Heep Hong Society, the biggest organization in Hong Kong that provides professional training and education to children with developmental and learning problems and their families, invited a group of experts consisting of psychologists, occupational therapists, special educators, and special child-care workers to translate the PEP-3 into Chinese (25) and conduct a validation study to examine the psychometric properties of the instrument. Based on a sample of 455 children with ASD and a comparison group of 281 children with typical development, the construct validity of the Chinese version of PEP-3 (CPEP-3) was examined. It was found that older children scored higher than younger children for both groups on different subsets of CPEP-3, while no gender difference was identified within the autistic group. The results of confirmatory factor analysis further showed the three-factor theoretical model of the Performance test fitted well. These findings provide important evidence for the construct validity of the CPEP-3 (25). However, besides construct validity, there is a need to further understand the reliability as well as other types of validity of the CPEP-3.

As far as reliability is concerned, three types of reliability are commonly assessed for an assessment tool (26). First, Cronbach's alphas are usually computed to examine the internal consistency of the scale to see whether the items measure something

which is homogenous in nature. Second, temporal stability of the test scores (i.e., test-retest reliability) is typically examined by getting the test scores at two different time points. Finally, if the test is completed by different raters, consistency across raters is examined by inter-rater reliability.

Regarding validity, besides construct validity which was examined in the study of Shek and Yu (25), two additional forms of validity evidence were examined in the present study. First, evidence about content validity was explored, which refers to "the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured" (27). Based on the method adopted by the original author of PEP-3 (12), conventional item analyses were performed to provide evidence for content validity, including item discrimination and item difficulty analyses. Second, criterion-related validity was investigated, which was defined as the correlation between a test and an (or a series of) external criterion variable(s) taken as representative of the construct, for example, measures that have previously been validated assessing presumably related constructs (27). In this study, it was hypothesized that the Performance Test of CPEP-3 should correlate well with previously validated tests that measure developmental skills and behavioral symptoms in children with autistic disorders.

The primary purpose of this paper is to report the reliability and validity of the Performance Test of the Chinese version of PEP-3 (CPEP-3) based on further data collected from the same samples of children as reported in Shek and Yu's (25) paper. Psychometric properties of the Caregiver Report will be presented elsewhere (28). Specifically, the current study attempted to assess the reliability (internal consistency, test-retest reliability and inter-rater reliability) and validity (content validity and criterion-related validity) of the CPEP-3 Performance Test as a measure to estimate the developmental level among children with ASD in Hong Kong.

Methods

Prior to the study, approval to conduct the study (including ethics approval) was obtained from the Executive Committee of Heep Hong Society. The

primary caregivers of the participated children have given their written informed consents to the research team. Participants were ensured that the data collected in the study would be kept confidential.

Participants

As described in Shek and Yu's (25) paper, 455 children who were diagnosed as having autism or other pervasive developmental disorders (PDDs) in 25 service units in Heep Hong Society including special child care centers, early education and training centers, and parent resource center were recruited to

participate in the study. Another sample of 281 children without developmental problems was selected as the "healthy control" from 13 local kindergartens matched for age with the autistic sample for comparison. The diagnoses of the autistic sample were made based on ICD-10/DSM-IV by consultant psychiatrists and endorsed by a multidisciplinary team consisting of clinical psychologists, special educators, and other helping professionals. Several subgroups of participants were randomly selected for the analyses of test-retest reliability, inter-rater reliability, and criterion-prediction validity.

Table 1. Demographic characteristic of the samples

Age group	2	3	4	5	6	7	Total
Age range in years	2.0-2.9	3.0-3.9	4.0-4.9	5.0-5.9	6.0-6.9	7.0-7.9	
Sample with autism							
No. of participants	32	79	140	161	37	6	455
Percentage	7.0%	17.4%	30.8%	35.4%	8.1%	1.3%	100%
No. of girls	6	11	22	22	5	0	66
No. of boys	26	68	118	139	32	6	389
Typically developing sample							
No. of participants	67	60	60	62	30	2	281
Percentage	23.8%	21.4%	21.4%	22.1%	10.7%	0.7%	100%
No. of girls	34	37	36	30	14	1	152
No. of boys	33	23	24	32	16	1	129

Participants' age ranges from 2.0 to 7.9 years. The characteristics of the two samples in terms of age and gender can be seen in Table 1. For the autistic sample, the ratio of boys to girls is 6:1, reflecting the fact that boys have a higher risk of autistic disorders than girls. While this figure is higher than the related ratio (4:1); in the sample reported in Schopler et al. (12), it is highly similar to the findings of a large epidemiological study of autistic spectrum disorder in which the male to female ratio was found to be 6.58:1 in Hong Kong children (8). Scores of the sample with ASD and the typically developing sample were compared as an indication of criterion-related validity.

Procedures

The translation of the CPEP-3 followed standard procedure in questionnaire adaptation, of which the details were reported in Shek and Yu's (25) paper. A group of professionals including speech therapists, occupational therapists, educational psychologists, and preschool teachers administered the tests and rated the participants' performance. All of the raters had experience working with and testing young children. Before the formal launch of the validation study, training sessions were conducted so that the raters worked together to clarify and get familiar with general testing, scoring and interpreting procedures

of the CPEP-3 to ensure consistency in the test administration.

To investigate whether the participants' test performance is consistent over time, test-retest reliability was examined on a subsample of 42 children with autism. Six girls and 36 boys between 2.3 and 5.9 years of age were tested twice over a period of time that ranged from 6 weeks to 3 months. The variation of the interval was due to difficulties in arrangement involved.

Besides, inter-rater reliability was conducted on a subsample of 46 children randomly selected from the autistic group with age ranging from 2 to 5 years (5 girls and 41 boys) to understand whether the test would be affected by subjective ratings of different raters. Each subject was administered the CPEP-3 Performance test independently by one administrator and two co-raters as observers. The administrator-co-raters grouping was randomly assigned by a research assistant who was blind to the characteristics of the subjects. In order to keep neutrality throughout the process, testing conditions were similar for all the children and the researchers involved in the case would rate the child's performance independently and handed the protocols to the research assistant directly for data entry.

To test the criterion-related validity of CPEP-3 Performance Test, two established measures related to developmental problems were administered on a subsample of 64 children with autism (56 boys and 8 girls) aged 2 to 6 years. The two measures included the Merrill-Palmer Revised Scales of Development (MPR) (29) and the Childhood Autism Rating Scale (CARS) (18). Both measures were administered by the raters and the participants were randomly assigned to the researchers. The correlations between the criterion measures and CPEP-3 Performance Test would serve as an index of criterion-related validity.

Instruments

The Chinese version of Psycho-Educational Profile—3rd Edition (CPEP-3)

The PEP-3 has two major parts: Performance Test and Caregiver Report. The 172-item Performance Test is composed of 10 subtests. Three subtests measure communication ability including Cognitive Verbal /

Preverbal (34 items), Expressive Language (25 items), and Receptive Language (19 items). Another three subtests measure motor ability: Fine Motor (20 items), Gross Motor (15 items), and Visual-Motor Imitation (10 items). These six subtests focus on the child's developmental level. The rest four subtests measure maladaptive behaviors, including Affective Expression (11 items), Social Reciprocity (12 items), Characteristic Motor Behaviors (15 items), and Characteristic Verbal Behaviors (11 items). The Caregiver Report consists of 38 items which are combined into three subtests: Problem Behavior (10 items), Personal Self-Care (13 items), and Adaptive Behavior (15 items). The current study focuses on examining the Performance section.

Authorized by the PEP-3 developers, Heep Hong Society organized a working group comprising clinical, developmental and educational psychologists to translate the PEP-3 items into Chinese. The translated draft was then reviewed and modified by the group after discussion. Compared with the English version of PEP-3, changes in the CPEP-3 were made in the areas of language and the use of stimuli. Adaptation and modifications were conducted taking into account the cultural and language factors. Chinese words were used to replace the English ones in the items for letter matching, naming and sorting, and a few more culturally suitable pictures were used to replace the original ones. The scoring of items has been quantified as 0, 1, and 2, with "Pass" = 2, "Emerge" = 1, and "Fail" = 0.

To assess the convergent validity of the CPEP-3, the following measures were employed:

Merrill-Palmer Revised Scales of Development (MPR)

The MPR is an individually-administered, norm-referenced measure of children's development in cognitive, language, motor, self-help, and social-emotional domains (29). It is characterized by a wide range of visual-motor tasks and a few verbal items. This instrument has been widely used by practitioners in assessing the cognitive functioning of pre-school children with language and cognitive delays, with high scores suggesting high cognitive functioning. There are seven subscales in the MPR: development, cognitive, fine motor, infant language/receptive language, infant memory/memory, speed, and visual

motor, measuring different areas of cognitive functioning. It was hypothesized that positive correlation would be found between MPR scores and CPEP-3 Performance subtest scores.

children with autism, Shek et al. (16) reported a significant correlation coefficient of 0.70 between the Behavioral Scale of CPEP-R and the total score of CARS.

Childhood Autism Rating Scale (CARS)

Data analyses

The CARS is a behavioral rating scale developed by the TEACCH Division (18). The CARS consists of ratings in 15 different areas of functioning significant for autism: Relating to People, Imitation, Emotional Response, Body Use, Object Use, Adaptation to Change, Visual Response, Listening Response, Taste, Smell, and Touch Response and Use, Fear or Nervousness, Verbal Communication, Nonverbal Communication, Activity Level, Level and Consistency of Intellectual Response and General Impressions. Children’s behaviors were observed and rated on a 7-point scale (4 anchor points and 3 half steps) with clear behavioral anchors. High scores on the CARS are considered more maladaptive. Both total score and subscale scores of the CARS were used in the current validity test. It is expected that negative correlation would be found between the CARS and the CPEP-3. On a sample of Hong Kong

Preliminary data screening was performed to verify the suitability of the current data for parametric analyses. The psychometric properties of CPEP-3 were examined at both the subtests and composite levels. First, three types of reliability tests were conducted to examine the consistency of CPEP-3 subtests and composites on the current samples of Hong Kong children. These included internal consistency, test-retest reliability, and inter-rater reliability. Second, conventional item analyses including both item difficulty and item discrimination were performed to provide evidence for the content validity. Third, convergent validity of CPEP-3 was examined by testing the correlation coefficients between CPEP-3 and the two criterion scales, MPR and CARS, which measure similar domains of development as measured by CPEP-3.

Table 2. Cronbach’s alpha coefficients and mean inter-item correlation coefficients for the CPEP-3 performance subtests and composites

Age Group	2	3	4	5	6	7	Average	Mean Inter-Item Correlation Coefficients	No. of Items
Performance Subtests									
CVP	0.92	0.96	0.96	0.97	0.93	0.60	0.97	0.48	34
EL	0.91	0.96	0.96	0.97	0.94	0.92	0.97	0.53	25
RL	0.92	0.96	0.96	0.97	0.92	0.60	0.97	0.62	19
FM	0.84	0.89	0.88	0.90	0.77	0.52	0.91	0.32	20
GM	0.82	0.88	0.86	0.89	0.62	-	0.90	0.38	15
VMI	0.85	0.90	0.88	0.88	0.71	-	0.90	0.48	10
AE	0.88	0.87	0.91	0.90	0.45	0.72	0.89	0.43	11
SR	0.86	0.86	0.89	0.91	0.81	0.87	0.90	0.45	12
CMB	0.89	0.92	0.91	0.92	0.89	-	0.92	0.41	15
CVB	0.94	0.95	0.93	0.95	0.84	0.64	0.95	0.62	11
Composites									
Communication	0.97	0.98	0.99	0.99	0.97	0.93	0.99	0.51	78
Motor	0.93	0.96	0.95	0.96	0.87	0.82	0.96	0.36	45
Maladaptive Behaviors	0.96	0.97	0.97	0.97	0.93	0.89	0.97	0.40	49

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

Results

Internal consistency was first examined for the subtests and composites of CPEP-3 Performance Test using the total autistic sample and the subsamples at six age intervals separately. The Cronbach's alpha coefficients and mean inter-item correlation coefficients are reported in Table 2. For the whole autistic sample, the Cronbach's alpha coefficients ranged from 0.89 to 0.97 for the Performance subtests and ranged from 0.96 to 0.99 for the composites. The values of the mean inter-item correlation coefficients were also acceptable (equalled to or exceeded 0.32

and 0.36 for the Performance subtests and the composites, respectively).

To investigate the temporal stability of the assessment tools, test-retest reliability was examined on a subsample of 42 randomly selected children with autism. Means and standard deviations of participants' scores at the two time points are presented in Table 3. Correlation coefficients between test and retest raw scores for each subtest were computed. Results showed that test-retest correlation coefficients were all above 0.82, indicating a good time sampling reliability of the Chinese version of PEP-3.

Table 3. Test-retest reliability for the CPEP-3 performance subtests

PEP-3 Values	First Testing		Second Testing		r
	M	SD	M	SD	
Performance Subtests					
CVP	38.26	16.89	40.00	17.36	.98
EL	13.07	11.26	13.79	12.30	.99
RL	18.00	11.43	18.76	11.43	.97
FM	31.19	6.46	31.62	6.66	.96
GM	25.38	5.26	26.29	5.20	.92
VMI	12.62	4.70	13.26	5.11	.93
AE	13.43	3.91	14.05	3.40	.84
SR	13.14	4.27	13.74	3.93	.92
CMB	20.71	6.91	21.40	6.69	.95
CVB	7.86	5.99	8.38	5.97	.96

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

Table 4. Polychoric correlation coefficients for CPEP-3 performance subtest items

Performance Subtests	Min	Max	Mean	Median
CVP	0.26	1.00	0.78	0.87
EL	0.41	1.00	0.77	0.81
RL	0.47	0.93	0.71	0.77
FM	0.23	0.96	0.60	0.76
GM	0.14	0.95	0.58	0.62
VMI	0.63	0.91	0.78	0.79
AE	0.11	0.59	0.34	0.34
SR	0.25	0.75	0.46	0.41
CMB	0.47	0.61	0.52	0.48
CVB	0.19	0.71	0.48	0.52

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

The reliability between raters was also assessed to understand whether the test would be affected by subjective ratings of different raters. The inter-rater reliability was based on a randomly selected subsample of 46 children with age ranging from 2 to 5 years (5 girls and 41 boys). Two experienced raters evaluated the 46 children independently using the 10 Performance subtests of CPEP-3. Polychoric correlation coefficients for each pair of items rated by different examiners were calculated as an indicator of inter-rater reliability (30). Descriptive statistics about the correlation coefficients for ten Performance subtests are summarized in Table 4. Because polychoric correlation requires bivariate normality of data, skewness of items was computed initially. Results revealed that 61 out of 344 items skewed greater than the absolute value of 2. Among these items, 12 items (items 42, 120, 155, 3, 70, 97, 56, 62, 69, 135, 137, and 138) yielded polychoric correlation coefficients that appeared to be significantly biased under the condition of bivariate normality. These correlation coefficients were then excluded in the calculation of the range, mean and median of the polychoric correlation coefficients for their corresponding subtests. As can be seen in Table 4, the mean and median polychoric correlation coefficient for each subtest ranged from 0.34 to 0.78 and from 0.34 to 0.87, respectively. According to Hopkin and Stanley (31), these values can be regarded as evidence showing good inter-rater reliability.

Validity of CPEP-3 Performance Test

Regarding the content validity, the rationales underlying the selection of subtest items were established by the test developers in the United States. In fact, an earlier version of PEP-R in Chinese was used in Hong Kong to assess children with autism (14, 32) and showed sufficient clinical validity (33). The PEP-3 is a modified version of the original PEP-R based on the experience of using this instrument in the United States for more than 20 years. After being translated into Chinese, all test materials and procedures of CPEP-3 were scrutinized by a steering group consisting of local experts including psychologists, occupational therapists and special teachers in Hong Kong. They generally held the expert view that items of the CPEP-3 are representative and highly relevant to the domains assessed.

Besides content validity supported by expert views, tests of item discrimination and item difficulty were conducted. Item discrimination refers to the extent to which success on an item corresponds to success on the whole test, which can be indexed by the correlation between a specific item and the other items of the questionnaire. Consistent with the method used by the original scale developers, item-total score correlation coefficients were calculated on the current autistic samples of children (age 2 through age 7) to represent for the item discrimination of the CPEP-3 items.

Table 5. Item discrimination for the CPEP-3 performance subtests

Performance Subtests	2	3	4	5	6	7	Overall
CVP	0.54	0.62	0.67	0.74	0.61	0.35	0.72
EL	0.63	0.72	0.77	0.81	0.71	0.85	0.83
RL	0.64	0.81	0.75	0.80	0.65	0.22	0.80
FM	0.46	0.58	0.59	0.63	0.48	0.77	0.61
GM	0.46	0.57	0.52	0.58	0.52	-	0.59
VMI	0.60	0.67	0.61	0.66	0.42	-0.33	0.68
AE	0.60	0.59	0.67	0.71	0.21	0.91	0.66
SR	0.56	0.56	0.61	0.66	0.51	0.72	0.63
CMB	0.61	0.63	0.58	0.61	0.59	0.11	0.64
CVB	0.83	0.78	0.73	0.77	0.52	0.60	0.76

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

Table 6. Item difficulty for the CPEP-3 performance subtests

Performance Subtests	2	3	4	5	6	7
CVP	18.8%	42.4%	59.3%	69.9%	89.2%	100%
EL	3.1%	19.0%	36.4%	49.7%	75.7%	83.3%
RL	21.9%	45.6%	62.9%	73.9%	89.2%	100%
FM	48.5%	70.3%	87.2%	91.0%	96.0%	100%
GM	53.1%	75.9%	88.6%	93.8%	100%	100%
VMI	14.1%	40.5%	62.5%	67.1%	82.5%	100%
AE	53.1%	46.8%	57.9%	68.3%	73.0%	83.3%
SR	21.9%	22.8%	44.3%	48.1%	63.6%	83.3%
CMB	43.8%	51.9%	64.3%	67.7%	73.0%	100%
CVB	12.5%	20.5%	35.7%	42.2%	64.9%	66.7%

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

Table 7. Descriptive information on two criterion measures and their subscales

		Mean	SD	Range
MPR	Development	179.27	54.86	38.00-237.00
	Cognitive	82.20	22.22	24.00-110.00
	Fine motor	47.20	12.27	13.00-64.00
	Infant language/receptive language	49.61	21.80	0.00-71.00
	Infant memory/memory	15.95	8.78	0.00-34.00
	Speed	14.72	4.70	2.00-20.00
	Visual motor	44.84	10.42	10.00-55.00
CARS	Relating to people	2.52	0.53	1.50-4.00
	Imitation	2.16	0.82	1.00-4.00
	Emotional response	1.82	0.68	1.00-3.50
	Body use	2.08	0.54	1.00-3.50
	Object use	2.30	0.50	1.00-3.50
	Adaptation to change	2.03	0.68	1.00-3.50
	Visual response	2.23	0.74	1.00-3.50
	Listening response	2.32	0.51	1.00-3.00
	Taste, smell, and touch response and use	1.78	0.68	1.00-3.00
	Fear or nervousness	1.24	0.62	1.00-4.00
	Verbal communication	2.59	0.75	1.00-4.00
	Nonverbal communication	1.95	0.62	1.00-3.00
	Activity level	1.52	0.65	1.00-3.00
	Level and consistency of intellectual response	1.84	0.62	1.00-4.00
	General impressions	2.69	0.63	1.00-4.00
Total score	31.05	6.03	21.50-46.00	

Table 5 shows the median item discrimination coefficients for each subtest of CPEP-3. Using the criterion that discrimination indexes of 0.35 or higher are acceptable (34-35), most items can be regarded as having good item-discrimination ability on both the whole sample and different age groups. For children aged at 7 years or older, item-discrimination indices for several subtests (RL, GM, VMI, CMB, and PSC) are either uninterpretable or less than 0.35, which may be due to the limited number of participants in this age group ($N = 6$).

Item difficulty refers to the percentage of participants who can pass a given item, which helps to identify whether an item is too difficult or too easy. Anastasi and Urbina (27) suggested an average item difficulty of 50%, with a range of 15% to 85% considered acceptable. Table 6 summarizes the median item difficulty for each subscale of CPEP-3 on the autistic sample of children. For most parts of the test, test items satisfy the requirements. Results showed that as age increased, the percentage of children who passed the item also increased. These results are also consistent with the Western findings (12).

To test the criterion-related validity of CPEP-3, two well-known measures, MPR and CARS were used as criterion measures based on a subsample of 64 randomly selected autistic children. Descriptive information about the two measures and their

subscales is reported in Table 7. As shown in Table 8, MPR subscale scores were positively correlated with the subtest scores of CPEP-3 Performance Test. All correlation coefficients were significant with values being moderate to large. This suggests high convergent validity of the Performance subtests of CPEP-3 when applied to Chinese children. For CARS, both its total score and subscale scores were used in the current validity test. As CARS measures malfunctioning in autistic children, it is expected that negative correlation would be found between the CARS and the CPEP-3 Performance subtests. As can be seen in Table 9, with a few exceptions, most correlation coefficients were significant, with the direction of the correlation coefficients consistent with the hypotheses, providing further evidence for the criterion-prediction validity of the CPEP-3. These findings were also comparable favorably with previous validation studies conducted in the United States using the CARS as one criterion measure for PEP-3 (12).

Finally, Table 10 summarizes the psychometric findings of CPEP-3 on the present sample of Chinese children and corresponding results of PEP-3 available in the PEP-3 user manual. It can be found that the psychometric properties of CPEP-3 compared favorably with those of PEP-3 reported by the test developers (12).

Table 8. Correlation coefficients between CPEP-3 performance subtests and MPR subscales

		Performance Subtests									
		CVP	EL	RL	FM	GM	VMI	AE	SR	CMB	CVB
MPR	Development	.93**	.92**	.91**	.81**	.72**	.79**	.70**	.80**	.73**	.81**
	Cognitive	.88**	.89**	.86**	.75**	.63**	.73**	.68**	.79**	.71**	.80**
	Fine motor	.92**	.87**	.88**	.81**	.75**	.80**	.65**	.72**	.65**	.75**
	Infant language/receptive language	.92**	.91**	.91**	.80**	.74**	.82**	.73**	.81**	.75**	.81**
	Infant memory/memory	.84**	.85**	.81**	.68**	.60**	.68**	.63**	.74**	.65**	.74**
	Speed	.95**	.89**	.90**	.85**	.78**	.80**	.66**	.74**	.66**	.76**
	Visual motor	.92**	.89**	.89**	.83**	.74**	.77**	.64**	.74**	.66**	.76**

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

** $p < .01$.

Table 9. Correlation coefficients between CPEP-3 performance subtests and CARS subscales

CARS subscales	Performance Subtests									
	CVP	EL	RL	FM	GM	VMI	AE	SR	CMB	CVB
Relating to people	-.55**	-.59**	-.65**	-.44**	-.42**	-.62**	-.63**	-.69**	-.74**	-.64**
Imitation	-.72**	-.72**	-.74**	-.66**	-.56**	-.72**	-.65**	-.76**	-.66**	-.73**
Emotional response	-.05	-.01	-.05	-.04	-.03	-.11	-.08	-.11	-.20	-.07
Body use	-.39**	-.44**	-.35**	-.32**	-.28*	-.28*	-.27*	-.43**	-.39**	-.32**
Object use	-.05	-.10	-.07	-.03	-.11	-.16	-.15	-.20	-.36**	-.07
Adaptation to change	.07	.08	.08	.12	.20	.01	-.01	-.07	-.11	.13
Visual response	-.24	-.31*	-.33**	-.17	-.17	-.25*	-.20	-.32**	-.34**	-.39**
Listening response	-.33**	-.36**	-.38**	-.28*	-.30*	-.30*	-.36**	-.43**	-.52**	-.40**
Taste, smell, touch response and use	-.37**	-.45**	-.40**	-.27*	-.23	-.39**	-.48**	-.59**	-.61**	-.50**
Fear or nervousness	.04	.02	.06	.08	.06	.00	.00	.19	.13	-.08
Verbal communication	-.67**	-.72**	-.75**	-.55**	-.47**	-.67**	-.68**	-.72**	-.77**	-.77**
Nonverbal communication	-.63**	-.71**	-.72**	-.49**	-.46**	-.46**	-.66**	-.70**	-.71**	-.76**
Activity level	-.49**	-.50**	-.47**	-.39**	-.31*	-.48**	-.56**	-.65**	-.66**	-.49**
Level& consistency of intellectual response	-.39**	-.43**	-.39**	-.27*	-.20	-.31*	-.53**	-.48**	-.51**	-.51**
General impressions	-.56**	-.62**	-.64**	-.43**	-.38**	-.57**	-.59**	-.70**	-.73**	-.64**
Total score	-.60**	-.65**	-.64**	-.46**	-.39**	-.60**	-.63**	-.77**	-.79**	-.68**

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors. ** $p < .01$; * $p < .05$.

Table 10. Comparison of the psychometric findings on CPEP-3 performance subtests and the findings in the PEP-3 manual

Performance Subtests	No. of Items	Average Cronbach's Alpha		Test-Retest Reliability		Median Inter-Rater Reliability		Correlation with CARS	
		Chinese (N = 455)	American (N = 407)	Chinese (N = 42)	American (N = 33)	Chinese (N = 46)	American (N.A.)	Chinese (N = 64)	American (N = 68)
CVP	34	0.97	0.97	0.98	0.99	0.87	-	-0.60	-0.75
EL	25	0.97	0.97	0.99	0.99	0.81	-	-0.65	-0.63
RL	19	0.97	0.97	0.97	0.99	0.77	-	-0.64	-0.76
FM	20	0.91	0.93	0.96	0.98	0.76	-	-0.46	-0.78
GM	15	0.90	0.92	0.92	0.99	0.62	-	-0.39	-0.70
VMI	10	0.90	0.95	0.93	0.97	0.79	-	-0.60	-0.78
AE	11	0.89	0.90	0.84	0.94	0.34	-	-0.63	-0.74
SR	12	0.90	0.91	0.92	0.97	0.41	-	-0.77	-0.68
CMB	15	0.92	0.91	0.95	0.98	0.48	-	-0.79	-0.71
CVB	11	0.95	0.93	0.96	0.95	0.52	-	-0.68	-0.65

Note. CVP = cognitive verbal/preverbal; EL = expressive language; RL = receptive language; FM = fine motor; GM = gross motor; VMI = visual-motor imitation; AE = affective expression; SR = social reciprocity; CMB = characteristic motor behaviors; CVB = characteristic verbal behaviors.

Findings of the American sample were excerpted from the PEP-3 manual (Schopler et al., 2005).

Discussion

The aim of the present study was to further report evidence on the reliability and validity of the Performance Test of the Chinese version of Psycho-Educational Profile-3rd Edition (CPEP-3) in the assessment of developmental disorders in a Chinese context. The findings showed that the Performance Test of the CPEP-3 has good psychometric properties when applied to Chinese population, which lends support to the cross-cultural applicability and stability of the CPEP-3.

There are several unique features of this study. First, this is a pioneering scientific study of the PEP-3 on a large sample of Chinese children who are under-represented in the scientific literature on ASD. Second, consistent with the approaches adopted in Schopler et al. (12), different types of reliability and validity were examined in this study. Third, the present findings are on par or even better than those reported in the original study. Finally, the sample size of the current study was quite large considering the relatively low incidence of ASD in children.

Three different types of reliability were examined. The internal consistency of each subtest measured by Cronbach's alpha coefficient for the whole sample of children (ranging from 0.89 to 0.97) was high in this study and is comparable with the findings (Cronbach's alpha coefficients ranging from 0.90 to 0.97) reported by Schopler et al. (12) on American samples. The values of the reliability measures appeared to be higher than the previous reports on the Chinese version of PEP-R (ranging from 0.74 to 0.98) in Hong Kong context (14), which suggests that the items of CPEP-3 are more homogeneous than CPEP-R in measuring their latent constructs. For the composites, their internal consistency was also high (ranging from 0.96 to 0.99). When the Cronbach's alpha coefficients were calculated for children at different ages, it was found that all values are high, except for participants of 7 to 7.9 years of age. This may be due to the limited number of cases ($N = 6$) at this age group and the lack of variance for a few items, which make it impossible to calculate any error variance.

For test-retest reliability, a subsample of children with autism was rated twice on the Performance test of CPEP-3 over six weeks to three months. The

correlation coefficients for each subtest at the two time points ranged from 0.84 to 0.99, suggesting high test-retest reliability. In other words, the participants' scores on CPEP-3 are unlikely to be influenced by random factors in the conditions of the participants or the testing environment over time. Inter-rater reliability was assessed for 46 children with ASD. According to the criterion suggested by Hopkin and Stanley (31) for determining the size of correlation coefficients ($0 < r < 0.1$ means very small correlation; $0.1 < r < 0.3$ means small correlation; $0.3 < r < 0.5$ means moderate correlation; $0.5 < r < 0.7$ means large correlation; and $0.7 < r < 0.9$ means very large correlation), the mean and median polychoric correlation coefficients for each Performance subtest ranged from moderate to very large. This provides support for the inter-rater reliability of the CPEP-3 Performance subtests. The findings suggest that using the CPEP-3, different examiners rated the items in a consistent way.

Consistent with western findings, traditional item analyses on the CPEP-3 Performance Test showed that both item discrimination and item difficulty were in an acceptable range, and the percentage of participants who passed a particular item also increased with age. These findings support the content validity of the instrument. The evaluation of the criterion-related validity as a measure to estimate the developmental level in comparison with two standardized developmental assessment measures (MPR and CARS) on a subsample of 64 autistic children demonstrated that CPEP-3 can be well used to effectively quantify the developmental level of children with ASD. Significant correlations were found between CPEP-3 and the two criterion measures, MPR and CARS. These findings echo with the findings reported by Schopler et al. (12) on samples of children in the United States and provide evidence for the criterion-related validity of CPEP-3.

Overall, the above findings support that the CPEP-3 Performance Test had good reliability and validity when applied to Hong Kong samples, which compared favorably with previous findings based on American samples. The summary shown in Table 10 clearly suggests that the translated CPEP-3 Performance Test is a reliable and valid measure for assessing Chinese children with autism and its results can be regarded as consistent.

It should be noted that there are several limitations of the present study. First, the study was conducted in Hong Kong and spoken Cantonese was used as the language for the administration of the test. Therefore, the participants may not be representative of the children with autism in a broader Chinese context, particularly for Mandarin-speaking areas. Further studies must be conducted in other Chinese societies and evaluate the psychometric properties of CPEP-3 when administered in Mandarin. Second, while one significant difference between CPEP-3 and PEP-R is the addition of a Caregiver Report, which rates children's performance based on caregivers' daily observation, the present study only investigated the psychometric properties of the Performance Test. As the adaptive behaviors and symptom severities of children with ASD vary across different environments, the Caregiver Report would provide important information on children's development in different contexts. To obtain a full picture of the development of children with ASD in Chinese contexts, reliability and validity of the Caregiver Report must be investigated. Third, although the group with autism and the control group were matched on participants' age in the present study, other factors that contribute to children's development were not taken into account, such as gender and IQ. Future study shall consider these factors when setting up the inclusion criterion in participants' recruitment. Nevertheless, as a promising assessment tool for children with ASD, the psychometric properties of CPEP-3 need to be established before they can be used to guide the development of intervention programs. Researchers and practitioners are most interested in obtaining scores that are accurate, reliable, and can best estimate the child's true level of development. The present study clearly suggests that the Chinese version of the Psycho-Educational Profile-3rd edition (CPEP-3) Performance Test has good reliability and validity when it is used to measure the developmental level of children with autism spectrum disorders in a Chinese context. In future, more evidence on the construct validity of the Performance test and the psychometric properties of the Caregiver Report of the CPEP-3 need to be further established.

Acknowledgments

The study was conducted by Heep Hong Society in Hong Kong. We thank colleagues of Heep Hong Society for collecting the data for the validation study. The authorship of the paper is equally shared by the first author and second author.

Ethical compliance

The authors have stated all possible conflicts of interest within this work. The authors have stated all sources of funding for this work. If this work involved human participants, informed consent was received from each individual. If this work involved human participants, it was conducted in accordance with the 1964 Declaration of Helsinki. If this work involved experiments with animals, it was conducted in accordance with the related institutions' research ethics guidelines.

References

- [1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-4. Washington, DC: American Psychiatric Association, 1994.
- [2] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Health Disorders: DSM-5. Washington, DC: American Psychiatric Publishing, 2013.
- [3] Rapin I, Katzman, R. Neurobiology of autism. *Ann Neurol* 1998;43:7-14.
- [4] Bertrand J, Mars A, Boyle C, Bove F, Yeargin-Allsopp M, Decoufle P. Prevalence of autism in a United States population: the Brick Township, New Jersey, investigation. *Pediatrics* 2001;108:1155-61.
- [5] Kielinen M, Linna SL, Moilanen I. Autism in Finland. *Eur Child Adolesc Psychiatry* 2000;9(3):162-7.
- [6] Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. *Res Autism Spectr Disord* 2011;5(1):418-25.
- [7] Scott FJ, Baron-Cohen S, Bolton P, Brayne C. Brief report: prevalence of autism spectrum conditions in children aged 5-11 years in Cambridgeshire, UK. *Autism* 2002;6:231-7.
- [8] Wong VCN, Hui SLH. Epidemiological study of autism spectrum disorder in China. *J Child Neurol* 2008;23:67-72.

- [9] Volkmar F, Cook EH, Pomeroy J, Realmuto G, Tanguay, P. Practice parameters for the assessment and treatment of children, adolescents, and adults with autism and other pervasive developmental disorders. American Academy of Child and Adolescent Psychiatry Working Group on Quality Issues. *J Am Acad Child Adolesc Psychiatry* 1999;38:32S-54S.
- [10] Salvia JA, Ysseldyke JE. Assessment in special and remedial education, 6th ed. Boston, MA: Houghton-Mifflin, 1995
- [11] Shriver MD, Allen KD, Mathews JR. Effective assessment of the shared and unique characteristics of children with autism. *Sch Psychol Rev* 1999;28:538-58.
- [12] Schopler E, Lansing MD, Reichler RJ, Marcus LM. Psychoeducational profile: TEACCH individualized psychoeducational assessment for children with autism spectrum disorders, 3rd ed. Austin, TX: Pro-Ed, 2005
- [13] De Leon C. Psychoeducational profile revised (PEP-R): the Brazilian version elaboration. *Autism* 2005; 9:450-1.
- [14] Portoghese C, Buttiglione M, Pavone F, Lozito V, De Giacomo A, Martinelli D, Margari Li. The usefulness of the revised psychoeducational profile for the assessment of preschool children with pervasive developmental disorders. *Autism* 2009;13:179-91.
- [15] Schopler E, Reichler RJ, Bashford A, Lansing MD, Marcus LM. Psychoeducational Profile-Revised (PEP-R), individualized assessment and treatment for autistic and developmentally disabled children, vol 1. Austin, TX: Pro-Ed, 1990.
- [16] Shek DTL, Tsang SKM, Lam LL, Tang FLY, Cheung, PMP. Psychometric properties of the Chinese version of the Psycho-Educational Profile-Revised (CPEP-R). *J Autism Dev Disord* 2005;35:37-44.
- [17] Villa S, Micheli E, Villa L, Pastore V, Crippa A, Molteni, M. Further empirical data on the psychoeducational profile-revised (PEP-R): reliability and validation with the Vineland adaptive behavior scales. *J Autism Dev Disord* 2001;40:334-41.
- [18] Schopler E, Reichler RJ, Renner BR. The Childhood Autism Rating Scale (CARS). Los Angeles, CA: Western Psychological Services, 1988.
- [19] Krug D, Arick J, Almond P. Autism Behavior Checklist—ABC. In: Krug DA, Arick J, Almond P, eds. Autism screening instrument for educational planning—ASIEP-2. Austin, TX: Pro-Ed, 1993.
- [20] Sparrow SS, Balla D, Cicchetti, DV. Vineland Adaptive Behavior Scales. Circle Pines, MN: American Guidance Services, 1984.
- [21] Kikas E, Häidkind P. Developing an Estonian version of the psychoeducational profile revised (PEP-R). *J Autism Dev Disord* 2003;33:217.
- [22] Chen KL, Chiang FM, Tseng MH, Fu CP, Hsieh, CL. Responsiveness of the Psychoeducational profile-third edition for children with autism spectrum disorders. *J Autism Dev Disord* 2011;41:1658-64.
- [23] Fu CP, Chen KL, Tseng MH, Chiang FM, Hsieh CL. Reliability and validity of the psychoeducational profile-third edition caregiver report in children with autism spectrum disorders. *Res Autism Spectr Disord* 2011;6:115-22.
- [24] Fu CP, Hsieh CL, Tseng MH, Chen YL, Huang WT, Wu PC, Chiang FM. Inter-rater reliability and smallest real difference of the Chinese Psychoeducational Profile-third edition for children with ASD. *Res Autism Spectr Disord* 2010;4:89-94.
- [25] Shek DTL, Yu L. Construct validity of the Chinese version of the Psycho-Educational Profile-3rd Edition (CPEP-3). *J Autism Dev Disord* 2014;44: 2832-43.
- [26] Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills, CA: Sage, 1979.
- [27] Anastasi A, Urbina S. Psychological testing, 7th ed. Upper Saddle River, NJ: Prentice Hall, 1997.
- [28] Shek DTL & Yu L. Psychometric properties of the Chinese version of the Psychoeducational Profile-Third Edition-Caregiver Report. *J Intellect Dev Disabil* 2015;40(4):321-29.
- [29] Roid GH, Sampers J. Merrill-Palmer Developmental Scale—Revised manual. Wood Dale, IL: Stoelting, 2004.
- [30] Jöreskog KG, Sörbom D. LISREL 7: A guide to the program and applications. Chicago, IL: SPSS, 1988.
- [31] Hopkins K, Stanley J. Educational and psychological measurement and evaluation, 6th ed. Englewood, NJ: Prentice Hall, 1981.
- [32] Lam MK, Rao N. Developing a Chinese version of the psychoeducational profile (CPEP) to assess children in Hong Kong. *J Autism Dev Disord* 1993;23: 273-9.
- [33] Tsang SKM, Shek DTL, Lam LL, Tang FLY, Cheung, PMP. Application of the TEACCH Program on Chinese pre-school children with autism—Does culture make a difference? *J Autism Dev Disord* 2007;37:390-6.
- [34] Ebel RL. Essentials of educational measurement, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1972
- [35] Pycszak F. Validity of the discrimination index as a measure of item validity. *J Educ Meas* 1973;10:227-31.

Submitted: September 10, 2016. *Revised:* September 28, 2016. *Accepted:* October 05, 2016.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.