

Validity and reliability of Preschool Language Scale 4 for measuring language development in children 48-59 months of age

Nuryani Sidarta^{*a}, Angela BM Tulaar^{**}, Amendi Nasution^{**} and Suryanto^{**}

ABSTRACT

*Department of Anatomy,
Medical Faculty,
Trisakti University
^{**}Department of Physical
Medicine and Rehabilitation
Cipto Mangunkusumo Hospital

Correspondence

^adr. Nuryani Sidarta, Sp.RM
Department of Anatomy,
Medical Faculty,
Trisakti University
Jl. Kyai Tapa 260 - Grogol
Jakarta 11440
Telp 021-5672731 ext.2101
Email: nuri_sid@yahoo.com

Univ Med 2008; 27: 174-82

Prevalence rates for speech and language delay have been reported across wide ranges. Speech and language delay affects 5% to 8% of preschool children, often persisting into the school years. A cross-sectional study was conducted in 208 children aged 48-59 months to determine the validity and reliability of the Indonesian edition of the Preschool Language Scale version 4 (PLS4) as a screening tool for the identification of language development disorders. Construct validity was examined by using Pearson correlation coefficient. Internal consistency was tested and repeated measurements were taken to establish the stability coefficient and intraclass correlation coefficients (ICC) for test-retest reliability. For construct validity, the Pearson correlation coefficient ranged from 0.151-0.526, indicating that all questions in this instrument were valid for measuring auditory comprehension (AC) and expressive communication skills (EC). Cronbach's alpha level ranged from 0.81-0.95 with standard error of measurement (SEM) ranging from 3.1-3.3. Stability coefficients ranged from 0.98-.0.99 with ICC coefficient ranging from 0.97-0.99 both of which showed an excellent reliability. This study found that PLS-4 is a valid and reliable instrument. It is easy to handle and can be recommended for assessing language development in children aged 48-59 months.

Keywords : Language development, PLS4, validity, reliability

INTRODUCTION

Language is a conventional code that has been socialized as a tool for expressing ideas or concepts, while speech is verbal communication for expressing language.⁽¹⁾ Language and speech

development has been named by the experts as an excellent indicator for assessing child development as a whole and cognitive ability and is associated with scholastic achievement.⁽²⁾ Rescola reported that the prevalence of language and speech disorders in children aged 2 - 4.5

years was around 5-8% and for language development disorders only it was around 2.3-19%.⁽³⁾ A number of studies even indicated that children below 5 years of age with language disorders that have not been treated correctly will have low verbal abilities, reading disorders and disorders of spelling and behavior.⁽⁴⁾ Roulstone (2003) found that children referred for speech therapy but not given direct intervention, within the next 12 months, two-thirds were still in the category of speech therapy.⁽⁵⁾ It is however a pity that to date there has been no clarification about the consistency of clinicians in screening for language and speech disorders in children.

One study stated that 43% of parents reported that their children (of age 10-35 months) never had been given assessment of development at the time their children visited the doctor for treatment and 30% of parents also reported that the pediatricians whom they visited had never discussed communication problems in their children.⁽⁶⁾ What formed an impediment in screening language and speech abilities of children usually was a time problem, absence of a definitive protocol and the fact that development problems are assumed to be less important than other complaints.

In Indonesia there are no complete data about the number of children experiencing language and speech disabilities. Rahmanoe (1999) reported a prevalence of 52.8% of children having delayed language abilities at the Growth and Development Clinic in Palembang.⁽⁷⁾ In Semarang, a study by Jaenudin found that the prevalence of children experiencing speech development problems was 4.7%.⁽⁸⁾ Data from the Department of Pediatrics, Cipto Mangunkusumo Hospital showed a value of 11.47%-12.9%.⁽⁹⁾

There are many assessments of linguistic and speech disorders in children, but there has never been uniformity in screening techniques. The instruments that are often used in evaluating

the communication skills in children are The Ages and Stages Questionnaire,⁽¹⁰⁾ Clinical Adaptive Test/Clinical Linguistic and Auditory Milestone Scale,⁽¹¹⁾ Denver Developmental Screening Test,⁽¹¹⁾ McArthur Communicative Developmental Inventory,⁽¹²⁾ Ward Infant Language Screening Test, Assessment, Acceleration, and Remediation (WILSTAAR),⁽¹³⁾ Fluharty Preschool Speech and Language Screening Test⁽¹¹⁾ and Early Language Milestone Scale.⁽¹¹⁾

The Preschool Language Scale (PLS) is a psychometric instrument that was designed to evaluate language ability in children from birth up to 6 years and 11 months^(14,15) This instrument is a diagnostic tool which may be applied for identification of comprehensive and expressive ability in children and may also be utilized to assess changes over time in the language abilities of the child. The Preschool Language Scale version 4 (PLS4) can be used individually and includes assessment of ability in preverbal behavior and at the same time assesses the language skills in the field of semantics, morphology, syntax, and integrative and preliterative abilities. The reliability and validity of PLS4 has been the subject of a number of studies.

Considering the increasing number of children with language and speech disorders and the absence of a gold standard for evaluation of these disorders, the present study aims to determine the validity and reliability of the Indonesian language edition of the PLS4 as an instrument for evaluating language development in Indonesian children.

METHODS

Research design

In this study an observational research design with cross-sectional approach was used for attaining the aims of this research.

Study subjects

Participants in this study were children 48-59 months of age from Kindergartens (nursery schools) in South Jakarta, and participation had been approved by the parents of the subjects. Non-cooperative children or those with abnormalities in vision and hearing and a history of neurological trauma in the central nervous system (CNS) were to be dropped out from the group of study participants. Respondents were recruited from four districts in South Jakarta, namely Tebet, Pasar Minggu, Cilandak and Mampang. From each district one Kindergarten was selected by cluster random sampling.

Sample size

The optimal sample size required for conducting this study was calculated on the basis of the prevalence of 16% of children with language disabilities, with an acceptable error of 0.05. Based on the results of the calculation, the minimum sample size was 208 children, with equal proportions of boys and girls.

Instrument

The PLS4 is an individually administered test designed to identify young children from birth to 6 years 11 months old who have a language disorder or delay. It consists of two subscales for the assessments of auditory comprehension and expressive comprehension respectively.⁽¹⁵⁾ The materials provided for the administration of the test include an examiner's manual that spells out the administration, scoring and interpretation procedures; a record form that provides all the items of the test; a picture manual that contains color pictorial stimuli needed for many of the test items; and a box of manipulatives (e.g., ball, bowls, rattles, windup toys, cloth) needed for interactions with the child during test

administration. It consists of 62 auditory comprehension tasks and 68 expressive comprehension tasks to identify comprehension and expressive language skills. However, almost all of these items/tasks are better described as testlets than items as each of them contains anywhere between 2 to 8 related sub-items/tasks.

Data analysis

In this study calculations were performed on internal and external reliability of PLS4. External reliability was assessed with the test-retest method, either by the same examiner (intraclass reliability) or by different examiners (interclass reliability). Internal reliability was obtained by calculating Cronbach's alpha where a minimum value of 0.7 was required to show reliability of an instrument.⁽¹⁵⁾ The validity measured in this study was construct validity, where for each question the validity coefficient was to be measured by finding a correlation between the score for each question and the total score obtained. The validity coefficient required had to be larger than 0.138 to demonstrate validity of a question.⁽¹⁵⁾ Construct validity was assessed by using Pearson's correlation coefficient for AC and EC. Internal and external validity was represented by Cronbach's alpha, the stability coefficient, and the intraclass correlation coefficient (ICC) with a time interval between examinations of 7 days.

RESULTS

A total of 208 participants was recruited during conduction of the study, consisting of 106 (51%) boys and 102 (49%) girls aged between 48-59 months, with a mean of 55.54 ± 2.86 months.

Table 1. Intraclass reliability of PLS4

	Test		Retest		Stability coefficient
	Mean	SD	Mean	SD	
Auditory comprehension	93.91	10.08	93.55	10.37	0.98
Expressive communication	87.82	10.96	87.73	10.96	0.99
Total language	90.64	10.19	89.95	10.41	0.98

Table 2. Interclass reliability value of PLS4

	Test		Retest		Interclass correlation
	Mean	SD	Mean	SD	
Auditory comprehension	93.55	10.37	93	10.37	0.97
Expressive communication	87.73	10.96	87.27	11.63	0.99
Total language	89.95	10.41	89.23	10.66	0.99

Reliability testing

The reliability of PLS4 in this study was to be assessed from the results of calculation of internal consistency (Cronbach's alpha), stability coefficient, and interclass correlation. With a total of respondents of 208 children, for the auditory comprehension domain an internal consistency was obtained of 0.95, with standard error of measurement (SEM) of 3.3. For the expressive communication domain a Cronbach's alpha of 0.92 with SEM 3.0 was obtained, while for total language domain a smaller internal consistency value of .081 with SEM 3.1 was obtained.

Table 1 below shows the intraclass reliability scores. Test and retest mean scores \pm SD for auditory comprehension were almost identical, being 93.91 ± 10.08 and 93.55 ± 10.37 , with a stability coefficient of 0.98. Test and retest mean scores for expressive communication were 87.27 ± 10.96 and 87.73 ± 10.96 , respectively, with a stability coefficient of 0.99, while mean total language scores were 90.64 ± 10.19 vs. 89.95 ± 10.41 , with a stability coefficient identical to that obtained for the auditory comprehension domain.

Table 3. Construct validity of auditory comprehension questions in PLS4

Question	Correlation (r)
36	0.196
37	0.188
38	0.298
39	0.291
40	0.264
41	0.187
42	0.361
43	0.396
44	0.359
45	0.347
46	0.306
47	0.234
48	0.449
49	0.349
50	0.366
51	0.403
52	0.360
53	0.388
54	0.407
55	0.518
56	0.382
57	0.151
58	0.196
59	0.188
60	0.298
61	0.291
62	0.264

Table 4. Construct validity of expressive communication questions of PLS4

Question	Correlation (r)
34	0.289
35	0.201
36	0.185
37	0.308
38	0.295
39	0.271
40	0.191
41	0.370
42	0.403
43	0.364
44	0.351
45	0.307
46	0.234
47	0.457
48	0.356
49	0.363
50	0.403
51	0.356
52	0.387
53	0.406
54	0.516
55	0.378
56	0.151
57	0.368
58	0.198
60	0.426
61	0.300
62	0.170

Table 5. Correlation between auditory comprehension, expressive communication and total language

	ACS	ECS	TLS
ACS	-	0.68	0.86
ECS	0.68	-	0.90
TLS	0.86	0.90	-

Legend: ACS (auditory comprehension standard score), ECS (expressive communication standard score), TLS (total language standard score)

To assess interclass reliability, test results were obtained by two different examiners with an interval of 7 days. Table 2 indicates that test and retest mean scores had an interclass correlation of 97% for the auditory comprehension domain and 99% for expressive communication and total language.

Results of analyses performed for each PLS4 question (Tables 3 and 4) indicated that the mean correlation for each question had an r value > 0.138.

Table 6. Distribution of values for auditory comprehension, expressive communication and total language in children aged 48-59 months with PLS4

Value	Minimum	Maximum	Median	Mean	SD	SEM
ACSS	65	125.0	99.0	97.8	9.6	3.3
AC-AE	32	75.0	55.0	55.7	6.3	3.5
ECSS	50	119.0	95.0	94.4	10.3	3.0
EC-AE	1	69.0	53.0	53.0	7.1	3.6
TLSS	50	115.0	96.0	95.8	10.5	3.1
TL-AE	11	69.0	53.0	54.1	7.0	3.7

Legend: ACSS (auditory comprehension standard score); AC-AE (auditory comprehension age equivalent); ECSS (expressive communication standard score); EC-AE (expressive communication age equivalent); TLSS (total language standard score); TL-AE (total language age equivalent); SD (standard deviation); SEM (standard error of measurement)

Correlation between domains

The correlation between the three language domains was also assessed in this study. The correlation between ACS and ECS was 0.68, while the correlation between ACS and TLS was 0.86 ($p < 0.05$). A larger correlation value was obtained in the association between ECS and TLS of 0.90 ($p < 0.05$) (Table 5).

Distribution of scores

In this study, the examiners also attempted to find the scores for auditory comprehension, expressive communication and total language. Table 6 shows a mean score of 97.8 ± 9.6 for auditory comprehension, with a median of 99, and a mean score of 94.4 ± 10.3 for the expressive communication domain, with a median of 95.0. For total language the mean score was 95.8 ± 10.5 , with a median of 96.0. For each domain the median approximated the mean, indicating that the distribution of scores in each domain was almost symmetrical, approximating the normal distribution.

DISCUSSION

The time required for completing the PLS4 test varied for each child, depending on whether or not the child was cooperative. In the present study of children aged 48-59 months, the time required ranged from 15-20 minutes. The results of this study differs from the assessment conducted by Zimmerman, where the time required ranged from 20-45 minutes.⁽¹⁴⁾ Screening instruments, which take about 10 minutes to administer, could offer a reasonable and standardized approach to screening for speech and language delay in primary care settings. However, screening with such a tool must be followed with a more thorough diagnostic evaluation before implementing an appropriate intervention.⁽¹⁵⁾

If the children were accompanied by a nursemaid or a parent, they usually would appear to be calm and could concentrate more on answering the questions. In addition, the number of children examined in a given room was also of influence on the time required for completion of the test. If the number of children was too large, the concentration of the children could be disturbed by voices of other children. On the other hand, if only one child was examined in a room, the child would be visibly fearful so that it had difficulties in answering the questions. According to the supplied reference book there is no standard size for examination rooms. The book only suggests that examination rooms are well-ventilated and have a background setting that is pleasing to the children, so that they feel comfortable when examined.

The results of this study show a Cronbach's alpha for components AC, EC, and TL respectively of 0.95, 0.92 and 0.81. A Cronbach's alpha larger than 0.7 demonstrates the high reliability of this instrument. In other words, this instrument has high degree of confidence, with a small random error. The results obtained in this study did not differ substantially from those of Zimmerman in his study on validation and reliability of PLS4 in 213 children, internal consistency values of 0.90, 0.93 and 0.95, respectively were obtained for AC, EC and TL.⁽¹⁴⁾ On average, the results of both studies were comparable, as both yielded high values for Cronbach's alpha, proving that the PLS4 instrument has a high reliability.

Cronbach's alpha indicates the level of consistency of the PLS4 instrument in measuring language ability, but cannot be directly used to interpret individual language skills. To this end the examiner also determined the SEM. This value will give information about the estimated deviations in language skills of a

given child from the test results obtained by that child. With high consistency values, the SEM may be expected to become smaller. In this study the SEM for AC was 3.3, while the SEMs for EC and TL were respectively 3.0 and 3.1. These values are slightly below the values obtained by Zimmerman⁽¹⁴⁾ in the validation study of PLS4, where the investigators obtained SEM values of 4.74, 3.97 and 3.21 for AC, EC and TL in the age group of 48-59 months. These SEM values will give estimates of the possible range of scores obtainable by a child. For example if a child gets a score of 9.0 with SD of 10 and SEM of 3, then the scores of the child range from 8.7 ± 10 up to 9.3 ± 10 , signifying that the scores of the child may vary from 7.7 to 10.3.

To test the external reliability of the PLS4 the examiner also calculated the stability coefficient for intraclass reliability and the inter-class correlation (ICC) for interclass reliability. The time interval used between the first and the second test was seven days. The reason for selecting this interval was because according to the manual the ideal time interval for the test and retest method is 2 to 14 days. The selected time interval was expected to minimize the effect of learning from the administered questions, but the interval was not too long to allow changes in language skills of the child.

In the present study mean test and retest scores for auditory comprehension were similar, namely 93.91 ± 10.08 and 93.55 ± 10.37 . The difference in mean scores between the first and second test was 0.36 with a difference in SD of 0.25 and stability coefficient of 0.98. This value indicates that if the assessment had been done twice by the same examiner, there would have been a probability of 98% of the results being identical. The values obtained here are lower than those in the reliability study by

Zimmerman⁽¹⁴⁾ in 218 children (117 girls and 101 boys) aged 2 to 5 year and 11 months, where the mean AC was 96.0 ± 15.1 and 98.1 ± 17.9 .⁽¹⁴⁾ The differences in mean values in the study by Zimmerman was 2.1, with difference in SD of 2.8 and stability coefficient of 0.85. Similar results were obtained for the two other domains. Mean test and retest scores for EC were 87.73 ± 10.96 and 87.27 ± 10.96 with stability coefficient of 0.99, while for TL they were 90.64 ± 10.19 and 89.95 ± 10.41 , with the same stability value as for auditory comprehension. In comparison, in the study by Zimmerman, obtained values for EC and TL for test and retest were 98.8 ± 15.0 and 101 ± 17.1 , respectively 97.2 ± 16.2 and 99.7 ± 18.5 with stability coefficient for AC, EC and TL of 0.95 and 0.96, respectively.

For assessment of inter-reliability, calculations of the ICC values were done by two different examiners on one subject. The time interval between examinations was seven days. To guarantee the objectivity of the assessments, the examiners did not know the scores given by one another. After completion of the scoring process the inter-reliability values were calculated. The obtained ICC for the AC domain was 0.974 and for EC and total language an ICC of 0.99 was found. This means that although tested by different persons, the probability of obtaining similar examination results was 97.4% with an error level of 2.6%. The same results were obtained in the study of PLS4 validation when testing the scoring system for this instrument in 100 samples. Every sample was assessed twice by different examiners and a correlation of 0.99 was found, indicating a high level of interclass reliability. This means that the scoring system in PLS4 was well-constructed and still had a good consistency even though examined by different persons.

After making statistical calculations for each of the questions found on the PLS4 (Tables 3 and 4) a range of correlation values was found for the auditory comprehension domain, from 0.17 to 0.52. For the expressive communication domain a correlation value of 0.151 to 0.516 was obtained, signifying that both the questions on auditory comprehension as well as those on expressive communication had a correlation value of more than 0.138. This indicates that abovementioned questions had construct validity, from which it may be concluded that the questions found in this instrument had formed a good construct or framework in assessing the language skills of children between 48-59 months of age.

To find an association between the respective domains the Pearson correlation test was performed. The correlation of ACS with ECS was 0.676, while the correlation of ACS with TLS was 0.856. A higher correlation was obtained for the association of ECS with TLS, being 0.903 with significant p value of < 0.05 . This indicates the existence of a substantially high correlation of ACS with ECS, because both domains are part of language and are interrelated. However, the correlation was not too strong because ACS and ECS are also two different aspects of language. On the other hand, the correlation of ACS-TLS and ECS-TLS had a higher correlation level, indicating that ACS and ECS scores would subsequently determine the total language score.

From the results in Table 6 it can be seen that the scores for all three domains has a normal distribution where the mean and medians are close together. In the auditory comprehension domain a range of 65 -125 was obtained, with mean score of 97.8 ± 9.6 . This value is slightly lower than that obtained in a study on 2000 children in the United States, where a mean score of 10.12 ± 13.8 was

obtained.⁽¹⁷⁾ However, if the normal range of scores is -1.5 SD to $+1.5$ SD, then the scores in Indonesian children are still within the normal range, namely between 77 and 122. Based on the above guidelines there were 12 children (58%) with scores below -1.5 SD. A study in Spain revealed that the children under study performed approximately 1.5 SD below the mean.⁽¹⁸⁾

For the domain of expressive communication the mean score obtained here was lower compared to the mean of African American children. In the present study the mean score was 94.4 ± 10.3 , while in African American children the obtained score was 101.7 ± 11.8 .⁽¹⁹⁾ However, similarly as with the AC scores, the EC scores in Indonesian children were still within the normal range. Results of statistical calculations indicated that there were 8 people (38%) with standard values below -1.5 SD. Mean total language score obtained in this study was 95.8 ± 10.5 and the number of children having a score below -1.5 SD was 25 (12%). However, this prevalence were not appreciably different from those obtained in the study by Elmina Harahap in children 3 to 24 months of age, where the prevalence of children experiencing language delay was 11.47%.⁽⁹⁾ PLS4 is frequently used for screening symptoms of neurodevelopmental delay (ND) and pervasive developmental disorder (PDD).⁽²⁰⁾

CONCLUSIONS

PLS4 is a valid instrument with a high level of reliability. This instrument is easy to administer and is recommended as one of the instruments for evaluation of language development in children aged 48-59 months. Overall, PLS4 is a well-designed and carefully developed instrument, meeting standards of quality.

REFERENCES

1. Owen R.E. The territory. In: Owen RE. Language development – nn introduction. Boston: Ally & Bacon; 2001. p. 12-29.
2. Catt HW, Fey ME, Tomblin JB. Zhang X. A longitudinal investigation of reading outcomes in children with language impairments. *J Speech Lang Hear Res* 2002; 45: 1142-7.
3. Rescorla L, Alley A. Validation of the language development survey (LDS): a parent report tool for identifying language delay in toddlers. *J Speech Lang Hear Res* 2001; 434-45.
4. US Preventive Services Task Force. Screening for speech and language delay in preschool children : recommendation statement. *Pediatrics* 2006;117 : 497-501.
5. Roulstone S, Peters TJ, Glogowska M, Enderby P. A 12-month follow up of preschool children investigating the natural history of speech and language delay. *Child Care Health Dev* 2003; 29: 245-55.
6. Halfon N, Olson I, Inkelas M, Lange I. Summary statistics from the National Survey of Early Childhood Health, 2000. *National Center for Health Statistics. Vital Health Stat* 2002; 15: 1-34.
7. Rahmanoe M. Spektrum faktor lingkungan yang berpengaruh terhadap terjadinya gangguan berbicara pada anak. *IKA-RSCM*; 1999.
8. Jaenudin E. Stimulasi keluarga pada perkembangan bicara anak usia 6 sampai 36 bulan di Kelurahan Kuningan, Semarang Utara (thesis). Semarang: IKA-/RSUP Dr Kariadi; 2000.
9. Harahap R.E. Perbandingan uji tapis Bayley Infant Neurodevelopmental Screener dengan uji tapis Denver II untuk deteksi dini keterlambatan perkembangan bayi usia 3-24 bulan dengan menggunakan baku emas Griffiths Mental Development Scales (thesis). Jakarta: IKA-RSCM; 2006.
10. Bricker D, Squires J, Mounts L. Ages and stages questionnaires: a parent-completed, child-monitoring system. 2nd ed. Baltimore: Paul H: Brookes; 1999.
11. Nelson HD, Nygren P, Walker M, Panoscha R. Screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics* 2006;117: e298-e319.
12. Fenson L, Pethick SJ, Renda C, Cox JL, Dale PS, Reznick JS. Short-form versions of the MacArthur Communicative Development Inventories. *Appl Psycholinguist* 2000; 21: 95–6.
13. Coulter L, Gallagher C. Piloting new ways of working: evaluation of the WILSTAAR Programme. *Int J Lang Commun Disord* 2001; 36: 270–5.
14. Zimmerman IL, Castilleja NF. The role of a language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities. Res Rev* 2005; 11: 238-46.
15. HOI KS. Review of the Preschool Language Scale – Fourth Edition, Pennsylvania State University, University Park, PA. Available at: <http://suen.ed.psu.edu/~hsuen/pubs/BUROS11%20preschool%20language%20scale.pdf>). Accessed September 3, 2008.
16. Nelson HD, Nygren P, Walker M, Panoscha R. Screening for Speech and Language Delay in Preschool Children. Rockville, MD: Agency for Healthcare Research and Quality. 2006. Evidence Synthesis No. 41. Available at: www.ahrq.gov/clinic/uspstfix.htm. Accessed September 3, 2008.
17. Zimmerman IL, Steiner VG, Pond RE. Preschool Language Scale, Fourth Edition (PLS-4). Technical Report. Available at: http://pearsonassess.com/NR/rdonlyres/6269B9FA-1CCF-4DF8-9F8B-282863589C25/0/PLS4_TechReport.pdf. Accessed August 30, 2008.
18. Restrepo MA, Silverman SW. Validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *Am J Speech Lang Pathol* 2001;10: 382-93.
19. Huaqing QI, Kaiser AP. The Performance of Low-Income African American Children on the Preschool Language Scale-3. *J Speech Lang Hear Res* 2003;46: 576-90.
20. Rescorla L. Language delay and behavioral/emotional problems in toddlers: findings from two developmental clinics. *J Speech Lang Hear Res* 2007; 50: 1063-78.