ORIGINAL ARTICLE

The Validity and Reliability Test of Modified Picture Vocabulary Test IV in 48-59 Months Old Children

Sumarni[1], Amendi Nasution[1], Luh KWahyuni[1], Trevino Pakasi[2]

[1] Department of Physical Medicine and Rehabilitation, Faculty of Medicine – University of Indonesia, Jakarta

[2] Department of Community Medicine, Faculty of Medicine – University of Indonesia, Jakarta

ABSTRACT

Objectives: To investigate the validity and reliability of Modified Picture Vocabulary Test IV (PPVT IV) in 48-59 months old child.

Methods 228 items in 19 sets of PPVT IV were translated in Indonesian language and revised according to Indonesian culture. One hundred and five 48-59 months old children were tested. The results were analyzed for its construct validity and reliability. The construct validity was valued by the correlation coefficient between the set score and the total score and the correlation coefficient between the item score and the set score. The reliability was valued by intra-class correlation coefficient, standard error of measurement (SEM) and Cronbach alpha.

Results: The construct validity of Modified PPVT IV was good ($r>0.256$; $\alpha=0.01$ $n=100$) and the reliability was high (internal consistency with Cronbach alpha= 0.938, inter-rater reliability= 0.957-0,985 and standard error of measurement (SEM) = 6.4).

Conclusions: Modified PPVT IV was a valid instrument with satisfactory reliability. To increase the construct validity, there were 36 items that should be substituted.

Key Words: *Modified PPVT IV, construct validity, internal consistency reliability, inter-rater reliability, standard error of measurement*

INTRODUCTION

The prevalence of speech and language development delay in children had been research excessively. The prevalence rate of speech and language development delay from abroad research was range from 5-8 %[1]. In Indonesia, the prevalence rate was range from 4.7-9.3%.[2, 3]

The language development delay issue is important because it will affect the child quality of life later on.[4] The result of the three longitudinal study showed that the language development was related to intelligence and reading ability afterwards. The language delay in early age is a strong predictor of low intelligence and reading difficulty 4-5 years later on. From that research, it is obvious that early intervention of children with speech and language delay is needed to prevent learning disability in school.

Received in March 2013 and accepted for published in April 2013.

Correspondence address: Sumarni, Jl. Tanjung Duren Barat III No. 26B, Jakarta Barat 11470. Email: sumarni_stephen@yahoo.com

Nowadays, in Indonesia there was no standardized test for evaluating the intervention success in one therapy for language delay. The standardized test is a language function evaluation test which is the most formal and decontextual.[5] The test was developed from research to a large amount of children with a normal language development and was counted for variation range that could be accepted. The benefit of this test, if made well, can be used to compare language development among children. The cause is that the standardized test has clear administration and scoring, good reliability and validity. One of the standardized test that is used in America is Peabody Picture Vocabulary Test (PPVT).[6]

This test had been modified in Mexican in 1976[7] and Spanish in 1986.[8] At this moment, there is no modified version in Indonesia. The researcher wanted to make the modified version in Indonesia according to Indonesian children culture and tested its validity and reliability. The PPVT IV itself has 2 forms, the Form A and the Form B. In this research, the Form A will be modified and tested for its validity and reliability.

METHODS

Research's samples were defined from selected population of samples using multi stage cluster random sampling. Inclusion criteria consist of all the kindergarten students which were 48-59 months old, was permitted by their parents to follow this research, and passed the training items.

Exclusion criteria consist of a student which wasn't permitted by their parents, was sick when the test was held, had been diagnosed by a doctor had mental retardation, speech and language delay, vision impairment that couldn't be corrected, hearing impairment, learning disability, emotional and behavioral disorder, ADHD (Attention Deficit Hyperactive Disorder) and genius.

After getting the ethical clearance from faculty ethical committee, the cross sectional study was started. 228 items in 19 sets of PPVT IV were translated in Indonesian language and were revised according to Indonesian culture. One hundred and five 48-59 months old children were tested. The results were analyzed for its construct validity and reliability. The construct validity was valued by the correlation between the item score and the set score. The reliability was valued by intra class correlation coefficient, Standard Error of Measurement (SEM) and Cronbach alpha.

RESULTS

There were 21 items in PPVT that were not suitable with Indonesian culture. They were revised according to Indonesian culture and were noticed for their difficulty level based on the child age and word category.

According to research characteristic of the subject, one hundred and five 48-59 months old children were tested. The reliability of modified PPVT IV was high (internal consistency) with Cronbach alpha = 0.938, inter-reliability = 0.957-0.985 and standard Error of Measurement (SEM) = 6.4. The construct validity of Modified PPVT IV was good ($r > 0.256; \alpha = 0.01; n = 100$).

Table 1. Characteristics of Study Samples

Socio-demography characteristics	n = 105
Age (Months)*	58 (48-59)
Gender **	
Male	48 (45.7)
Female	57 (54.3)
History of Chronic Suppurative Otitis Media**	
No	93 (88.6)
Yes	12 (11.4)
Mothers Education Level **	
Average (Graduated from Senior High School)	40 (38.1)
High (Academy 1-3 years)	56 (53.3)
Very High (Academy > 4 years)	9 (8.6)
Bilingualism**	
No	88 (83.8)
Yes	17 (16.2)
Family Income **	
Low (< 1 million/month)	5 (4.8)
Average (1-5 million/month)	77 (73.3)
High (> 5 million/month)	23 (21.9)

* = median (interquartil range), ** n(%)

DISCUSSION

In this modified PPVT IV, there were 21 items that were revised from its direct translation. In Mexican version there was no item that was modified from its direct translation.[7] The modification was just made for the arrangement of the items according to the difficulty level. In Spanish version, 50 items were deleted because of the culture bias.[8]

The amount of children that were tested for this research is one hundred and five. Compare with estimated sample sum, which is one hundred, the sample in this research is over 5 %. If compared with PPVT IV research in America in 2004 (200 children),[5] the sample amount of this research just reach 52.5% from that research. However, if compared with the research in Mexico, with 53 children, the sample amount of this research is quite lot.[7]

The Cronbach Alpha for Modified PPVT was 0.938. If compared with PPVT research in America, the Cronbach alpha for this version was lower (r = 0.965),[10] but it was almost similar with Spanish version (Cronbach alpha = 0.930)[11] and higher than Mexican version (Cronbach alpha = 0.85).[7] However, the Cronbach alpha value was more than 0.90, so the reliability is high.[12, 13]

Inter-rater reliability of this research was range from 0.957-0.985. This reliability is higher than PPVT in America (0.93).[14] This might be caused by adequate training and the good comprehension of the assistant.

The Standard Error of Measurement (SEM) in this research (6.4) is higher than research in America (3.3).[14] It means the reliability of American PPVT is higher than its modification version.

The validity of Modified PPVT IV to measure verbal ability in this research was done by testing its construct validity. The construct validity was measured by looking at the correlation between each set score with total score of PPVT. Now, in Indonesia, there is no golden standard for measuring child verbal ability, so according to literature, it was taken general value as validity limit.[15, 16] In this research, all r value in each set was more than 0.256 ($\alpha = 0.01$; n = 100). It means that all the sets of Modified PPVT IV can be used in determining PPVT IV score.

From validity test per set, there were 36 items that were not valid, (in set

1,2,3,4,5,6,7,8,11,18, and 19). That items were not valid because the low variability, so that the correlation with total score were low, too. For example, the items 1,2,3,4,5,6,9,10,11 (appendix table 1); and 13,19 (appendix table 2), in which the all respondents (100%) answered the items correctly, so the presence of that items did not affluence the set score.

The other items that were not valid or had low positive correlation showed that a child that answered the item correctly, this set score was low. So on the contrary, a child that answered the item incorrectly, this set score was high. The causes might be that the items confused the child, the drawing was not good so made misinterpretation, the item had no clear answer so the distracter had the chance to be chosen or was qualified as the correct answer, the item had distracter that almost children felt it was definitely right so the children guessed it was the right answer, there was gender, ethnic or subgroup bias in items or distracters.[17] In this research, the researchers didn't record the answered items (1, 2, 3 or 4), nevertheless right or wrong answers. That is why it could not be analyzed further about the reason that was the most possible. But for increasing the validity of each set, the items that were not valid could be changed.

From 36 items that was not valid, there were 6 items (16.67%) that were from revision output. The other items (83.33%) were from direct translation. Thus, the big part of the revision outcome was valid.

The reliability and validity was quite high in this research, so it was proved that a language test could be translated. It can spare the energy, time and cost if compared with make a new test tool.[7] This also suitable with "language universals" by Chomsky (1981) which stated that the languages in this world have the same foundation. The reason is that the language is correlated with human internal structure and human is the only one creature which has and uses language. So, the language is also based on universal concept based on neurology and biology similarity.[18]

The limitation of this research is that it just tested construct validity and internal consistency, inter-rater reliability and SEM of Modified PPVT IV. There are another validity and reliability that had not been tested. The validity that had not been tested is predictive, content and external validity. The reliability that had not been tested is intra-rater test-retest, equivalent instrument test and consolidation test.

Besides, in this research, the respondents are more in 54-59 months old, so it did not make a normal age distribution.

Another limitation is that the modified picture is a photo, not a painting like the original version. This make this modified PPVT cannot be used in color blindness children.

CONCLUSIONS

Modified PPVT IV was a valid instrument with satisfactory reliability. To increase the construct validity, there were 36 items that should be substituted.

REFERENCES

1. Nelson H, 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;117:298-319.
2. Jaenudin E. Stimulasi keluarga pada perkembangan bicara anak usia 6-36 bulan di Kelurahan Kuningan, Semarang Utara [Thesis]. Semarang: UNS; 2000.
3. Wahyuni S. Pemeriksaan penyaring keterlambatan perkembangan bahasa pada anak balita dengan Early Language Milestone Scale di Kelurahan Paseban Jakarta Pusat [Thesis]. Jakarta: FKUI; 1998.
4. Silva P. Epidemiology, longitudinal course and some associated factors : an update. In: Yule W, Rutter M, editors. Language Development and Disorders. Oxford: Mac Keith Press, Blackwell Scientific Publication Ltd; 1987.p.1-15.
5. Paul R. Evaluation and Assesment. In: Paul R, editor. Language Disorders from Infancy through Adolescence. 3rd ed. Philadelphia: Saunders Elsevier; 2007. p.41.
6. Paul R. Assesment of developing language. In: Paul R, editor. Language Disorders from Infancy through Adolescence. 3rd

ed. Philadelphia: Saunders Elsevier; 2007. p.322-8.
7. Simon A, Joiner L. A Mexican Version of The Peabody Picture Vocabulary Test. Journal of Educational Measurement. 1976;13:137-43.
8. Dunn L, Padilla E, Lugo D. Test de Vocabulario en Imagenes Peabody-Adaptation Hispanoamericana. American Guidance Service; Circle Pines, MN: 1986. Peabody Picture Vocabulary Test- Latin American adaptation.
9. Riduwan, Sunarto. Validitas dan Reliabilitas. In: Akdon, editor. Pengantar Statistika untuk Penelitian Pendidikan, Sosial, Ekonomi, Komunikasi dan Bisnis. Bandung: CV Alfabeta; 2007. p.347-60.
10. Dunn L, Dunn D. Content development, tryout, standarization and norms development. In: Dunn L, Dunn D, editors. PPVT4 Peabody Picture Vocabulary Test Manual. Minneapolis: NCS Pearson, Inc; 2007. p.27-30.
11. Moeller M. Intervention and language development in children who are deaf and hard of hearing. Pediatrics. 2000;106:1-9.
12. Sugiyono. Statistik untuk pengujian validitas dan reliabilitas instrumen penelitian. In: Mulyatiningsih E, editor. Statistika untuk Penelitian. 12th ed. Bandung: CV Alfabeta; 2007. p.348-66.
13. Tumbelaka A, Riono P, Sastroasmoro S, Wirjodiarjo M, Pudjiastuti P, Firman K. Pemilihan Uji Hipotesis. In: Sastroasmoro S, Ismael S, editors. Dasar-dasar Metodologi Penelitian Klinis. 2nd ed. Jakarta: CV Sagung Seto; 2002. p.254.
14. Dunn L, Dunn D. Technical Characteristics. In: Dunn L, Dunn D, editors. PPVT 4 Peabody Picture Vocabulary Test. Minneapolis: NCS Pearson, Inc; 2007. p.53-63.
15. Choudhury N, Benasich A. A family aggregation study: the influence of family history and other risk factors on language development. Journal Speech Language Research. 2003;46:261-72.
16. Paul R. Assessment and intervention for emerging language. In: Paul R, editor. Language Disorders from Infancy through Adolescence. 3rd ed. Philadelphia: Saunders Elsevier; 2007. p.266-7.
17. Varma S. Preliminary Item Statistics Using Point-Biserial Correlation and P-Values [Internet]. 2008. Available from: http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf
18. Sidiarto L. Berbagai gangguan berbahasa pada anak. In: Dardjowidojojo S, editor. PELLBA 4 Pertemuan Linguistik Lembaga Bahasa Atma Jaya. Jakarta: Kanisius; 1991. p.135.

APPENDIX

Table 1. The Validity of Each Item in the 1st Set Compared with the 1st Set Score

Item	r_{count}	Decision
1	0.000	Not valid
2	0.000	Not valid
3	0.000	Not valid
4	0.000	Not valid
5	0.000	Not valid
6	0.000	Not valid
7	-0.047	Not valid
8	0.241	Valid
9	0.000	Not valid
10	0.000	Not valid
11	0.000	Not valid
12	0.355	Valid

$r_{Table} > 0.195$ is is assumed valid ($\alpha = 0.05$; $n = 100$)[9]

Table 2. The Validity of Each Item in the 2nd Set Compared with the 2nd Set Score

Item	r_{count}	Decision
13	0.000	Not valid
14	0.149	Not valid
15	0.332	Valid
16	0.177	Not valid
17	0.186	Not valid
18	0.209	Valid
19	0.000	Not valid
20	0.044	Not valid
21	0.148	Not valid
22	0.093	Not valid
23	0.183	Not valid
24	0.092	Not valid

r_{Table} > 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 3. The Validity of Each Item in the 3rd Set Compared with the 3rd Set Score

Item	r_{count}	Decision
25	0.542	Valid
26	0.340	Valid
27	0.215	Valid
28	0.054	Not valid
29	-0.100	Not valid
30	0.358	Valid
31	0.134	Not valid
32	0.427	Valid
33	0.245	Valid
34	0.019	Not valid
35	0.497	Valid
36	0.270	Valid

r_{Table} > 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 4. The Validity of Each Item in the 4th Set Compared with the 4th Set Score

Item	r_{count}	Decision
37	0.020	Not valid
38	0.476	Valid
39	0.643	Valid
40	0.233	Valid
41	0.261	Valid
42	0.088	Not valid
43	0.271	Valid
44	0.621	Valid
45	-0.035	Valid
46	0.277	Valid

47	0.643	Valid
48	0.201	Valid

r_{Table} > 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 5. The Validity of Each Item in the 5th Set Compared with the 5th Set Score

Item	r_{count}	Decision
49	0.455	Valid
50	0.419	Valid
51	0.304	Valid
52	0.336	Valid
53	0.346	Valid
54	0.296	Valid
55	0.246	Valid
56	0.419	Valid
57	0.510	Valid
58	0.244	Valid
59	0.231	Valid
60	0.253	Valid

r_{Table} > 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 6. The Validity of Each Item in the 6th Set Compared with the 6th Set Score

Item	r_{count}	Decision
61	0.325	Valid
62	0.183	Not valid
63	0.209	Valid
64	0.201	Valid
65	0.200	Valid
66	0.273	Valid
67	0.313	Valid
68	0.203	Valid
69	0.385	Valid
70	0.175	Not valid
71	0.374	Valid
72	0.375	Valid

r_{Table} > 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 7. The Validity of Each Item in the 7th Set Compared with the 7th Set Score

Item	r_{count}	Decision
73	0.341	Valid
74	0.222	Valid
75	0.536	Valid

76	0.535	Valid
77	0.443	Valid
78	0.165	Not valid
79	0.112	Not valid
80	0.172	Not valid
81	0,420	Valid
82	0.273	Valid
83	0.587	Valid
84	0.417	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 8. The Validity of Each Item in the 8th Set Compared with the 8th Set Score

Item	r_{count}	Decision
85	0.835	Valid
86	0.324	Valid
87	0.440	Valid
88	0.514	Valid
89	0.435	Valid
90	0.675	Valid
91	0.488	Valid
92	0.422	Valid
93	0.639	Valid
94	0.529	Valid
95	0.338	Valid
96	0.170	Not valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 9. The Validity of Each Item in the 9th Set Compared with the 9th Set Score

Item	r_{count}	Decision
97	0.300	Valid
98	0.788	Valid
99	0.547	Valid
100	0.339	Valid
101	0.850	Valid
102	0.880	Valid
103	0.369	Valid
104	0.769	Valid
105	0.220	Valid
106	0.224	Valid
107	0.305	Valid
108	0.536	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 10. The Validity of Each Item in the 10th Set Compared with the 10th Set Score

Item	r_{count}	Decision
109	0.794	Valid
110	0.551	Valid
111	0.271	Valid
112	0.559	Valid
113	0.411	Valid
114	0.624	Valid
115	0.356	Valid
116	0.520	Valid
117	0.759	Valid
118	0.276	Valid
119	0.225	Valid
120	0.667	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 11. The Validity of Each Item in the 11th Set Compared with the 11th Set Score

Item	r_{count}	Decision
121	0.483	Valid
122	0.411	Valid
123	0.568	Valid
124	0.701	Valid
125	0.940	Valid
126	0.560	Valid
127	0.672	Valid
128	0.940	Valid
129	0.506	Valid
130	0.471	Valid
131	0.898	Valid
132	0.181	Not valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 12. The Validity of Each Item in the 12th Set Compared with the 12th Set Score

Item	r_{count}	Decision
133	0.618	Valid
134	0.634	Valid
135	0.454	Valid
136	0.375	Valid
137	0.838	Valid
138	0.863	Valid
139	0.370	Valid
140	0.356	Valid

141	0.718	Valid
142	0.536	Valid
143	0.411	Valid
144	0.832	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 13. The Validity of Each Item in the 13th Set Compared with the 13th Set Score

Item	r_{count}	Decision
145	0.902	Valid
146	0.743	Valid
147	0.882	Valid
148	0.837	Valid
149	0.548	Valid
150	0.618	Valid
151	0.894	Valid
152	0.723	Valid
153	0.743	Valid
154	0.459	Valid
155	0.503	Valid
156	0.665	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 14. The Validity of Each Item in the 14th Set Compared with the 14th Set Score

Item	r_{count}	Decision
157	0.234	Valid
158	0.453	Valid
159	0.587	Valid
160	0.630	Valid
161	0.230	Valid
162	0.595	Valid
163	0.512	Valid
164	0.339	Valid
165	0.382	Valid
166	0.486	Valid
167	0.795	Valid
168	0.855	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 15. The Validity of Each Item in the 15th Set Compared with the 15th Set Score

Item	r_{count}	Decision
169	0.599	Valid
170	0.670	Valid
171	0.795	Valid
172	0.686	Valid
173	0.408	Valid
174	0.497	Valid
175	0.448	Valid
176	0.787	Valid
177	0.917	Valid
178	0.507	Valid
179	0.703	Valid
180	0.940	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 16. The Validity of Each Item in the 16th Set Compared with the 16th Set Score

Item	r_{count}	Decision
181	0.397	Valid
182	0.974	Valid
183	0.817	Valid
184	0.937	Valid
185	0.473	Valid
186	0.974	Valid
187	0.974	Valid
188	0.718	Valid
189	0.620	Valid
190	0.800	Valid
191	0.332	Valid
192	0.599	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 17. The Validity of Each Item in the 17th Set Compared with the 17th Set Score

Item	r_{count}	Decision
193	0.353	Valid
194	0.894	Valid
195	0.481	Valid
196	0.737	Valid
197	0.831	Valid
198	0.755	Valid
199	0.813	Valid
200	0.902	Valid

201	0.578	Valid
202	0.678	Valid
203	0.227	Valid
204	0.971	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 18. The Validity of Each Item in the 18th Set Compared with the 18th Set Score

Item	r_{count}	Decision
205	0.837	Valid
206	0.638	Valid
207	0.493	Valid
208	0.936	Valid
209	0.364	Valid
210	0.903	Valid
211	0.194	Not valid
212	0.492	Valid
213	0.449	Valid
214	0.194	Not valid
215	0.529	Valid
216	0.929	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]

Table 19. The Validity of Each Item in the 19th Set Compared with the 19th Set Score

Item	r_{count}	Decision
217	0.254	Valid
218	0.975	Valid
219	0.707	Valid
220	0.677	Valid
221	0.774	Valid
222	0.000	Not valid
223	0.774	Valid
224	0.975	Valid
225	0.869	Valid
226	0.817	Valid
227	0.975	Valid
228	0.426	Valid

r_{Table}> 0.195 is is assumed valid (α = 0.05; n = 100)[9]