

Reliability of the Indonesian version of the School-Years Screening Test for Evaluation of Mental Status-Revised as a cognitive screening tool for children

Ferriandis Harsono, Purboyo Solek, Kusnandi Rusmil

Abstract

Background Developmental and behavioral problems are among the most common conditions of childhood. These problems affect 12–16% of children in the United States and 13–18% of children in Indonesia. Early detection of developmental deficits among children requires clinicians to screen with accurate tools. Cognitive function screening in children has been increasingly used in many clinical and educational settings. The School-Years Screening Test for Evaluation of Mental Status-Revised (SYSTEMS-R) is becoming a widely-used, reliable, and valid cognitive screening tool for children aged 4–15 years. Prior to our study, there was no Indonesian language version of the SYSTEMS-R.

Objective To determine the reliability of the Indonesian translation of the SYSTEMS-R.

Methods A cross-sectional study was conducted in children aged 4–15 years who understood the Indonesian language and did not have neurologic or communication impairments. Data were analysed to determine reliability (internal consistency and inter-rater reliability) of the Indonesian version of SYSTEMS-R. Internal consistency was determined using Cronbach's alpha formula. Internal consistency is a reflection of inter-item correlation and item-to-total correlation. Inter-rater reliability was determined using the Bland-Altman method.

Results This study was conducted on 133 children aged 4–15 years in a kindergarten, elementary, junior high, and senior high school in Bandung. The Indonesian version of SYSTEMS-R had significant internal consistency (Cronbach's alpha 0.936–0.941), and the scores obtained by two raters had good agreement (difference within mean \pm 1.96 SD).

Conclusion The Indonesian version of SYSTEMS-R is reliable for use as a cognitive screening tool for Indonesian children. [Paediatr Indones. 2016;56:149-54].

Keywords: cognitive screening, SYSTEMS-R, Indonesia, reliability

Developmental and behavioral problems are the most common conditions of childhood and adolescence.¹ Developmental and behavioral problems accounted for 12–16% of children in the US^{2–4} and 13–18% of children in Indonesia.² A study reported that in 423 elementary school students in Jakarta, 24.6% of the students had learning disabilities.⁵ In addition, a study in Brazil showed that the prevalence of learning disabilities in first- and second-graders of a state-run school was 10%, with a 3:1 male-to-female ratio.⁶ Furthermore, in two US studies, the prevalence of learning disabilities throughout elementary education ranged from 17 to 27%.^{6–8}

Early detection of developmental deficits among very young children typically requires clinicians to

From the Department of Child Health, Universitas Padjadjaran Medical School/Dr. Hasan Sadikin General Hospital.

Reprint requests to: Ferriandis Harsono, Department of Child Health, Padjadjaran University Medical School/Dr. Hasan Sadikin Hospital, Jl. Pasteur No. 38, Bandung 40161, Indonesia, Tel. +62 22-2035957; E-mail: tiffyandis@yahoo.com.

Further use of School Years Screening Test for Evaluation of Mental Status-Revised (SYSTEMS-R) or the Indonesian Version of SYSTEMS-R requires permission from the copyright owners of SYSTEMS-R: Dr. Laurel Fisher (laurel.fisher@y7mail.com), Dr. Fiona Spencer (fiona167.spencer@yahoo.com.au), and Professor Robert Ouvrier (robert.ouvrier@health.nsw.gov.au or roberto@chw.edu.au).

screen with tools proven to be accurate. To improve detection in primary care, the American Academy of Pediatrics recommends developmental screening and surveillance at well visits.¹ At present, the School-Years Screening Test for the Evaluation of Mental Status-Revised (SYSTEMS-R) is becoming a widely-used, reliable, and valid cognitive-screening tool for children aged 4–15 years.^{9–15} Prior to our study, with Indonesian as the primary language of the country, there was no Indonesian version of SYSTEMS-R. Some words in spelling items in the SYSTEMS-R needed adaptation to the Indonesian language and culture. For example, the word “cat” was not translated to “*kucing*,” but to “*ibu*” which had a similar difficulty level as “cat.” In addition, some proverbs in SYSTEMS-R also needed adaptation to Indonesian culture. For example, “*nasi sudah menjadi bubur*” was substituted for “don’t cry over spilt milk.”

Measurement instruments used in research, such as questionnaires, must be tested and proven to be reliable and valid.¹⁶ Reliability can be tested by measuring internal consistency, inter-rater reliability, or test-retest reliability. Validity can be tested by measuring construct validity, content validity, or criterion-related validity.^{17,18} Criterion-related validity considers whether scores on the instrument agree with a definitive “gold standard” measurement of the same theme.¹⁹ We determine how good a test is by comparing the test results with known diagnostic findings obtained by a reference standard. The reference standard reflects the patient’s true status. The validity of a diagnostic or screening tool is evaluated in terms of its ability to accurately assess the presence and absence of the target condition. Sensitivity is the test’s ability to obtain a positive test when the target condition is really present. Specificity is the test’s ability to obtain a negative test when the condition is really absent.²⁰

For the above background, we measured the reliability (internal consistency and inter-rater reliability) of the Indonesian translation of SYSTEMS-R. Due to time and fund constraints, measurements of other aspects of reliability (i.e., test-retest reliability) and validity (i.e., criterion-related validity, sensitivity, and specificity) of the Indonesian version of SYSTEMS-R were not yet performed.

Methods

Subjects were children aged 4–15 years at St. Paul Christian Kindergarten, Elementary, Junior High, and Senior High School, Bandung, West Java. Subjects (for children ≥ 12 years) or parents/representatives gave written informed consent. The inclusion criteria were children aged 4–15 years who understood the Indonesian language. The exclusion criteria were children with neurologic or communication impairments. This study was carried out in August 2014 with the approval of the Health Research Ethics Committee of the Padjadjaran University Faculty of Medicine /Dr. Hasan Sadikin Hospital.

This study had a cross-sectional design. For internal consistency, using a coefficient of correlation (r) of 0.6, 95% confidence interval (CI), and 80% power test, the minimum required sample size was 19 subjects per age group. For inter-rater reliability, using $r = 0.6$, 95% CI, and 95% power test, the minimum required sample size was 30. Subjects were chosen by stratified random sampling.

After obtaining permission from the copyright owners of SYSTEMS-R to do this study, SYSTEMS-R was translated to Indonesian by one translator, then translated back to English by a different translator. SYSTEMS-R consists of 40 items and uses a dichotomy scoring system, i.e. score 1 for a correct answer or score 0 for an incorrect answer. The research advisors reviewed, evaluated, and made corrections to the Indonesian version of SYSTEMS-R. After approval by the research advisors (subspecialists in Growth, Development, and Social Pediatrics and Pediatric Neurology), the Indonesian version of SYSTEMS-R was used in this study. Administration of the test was done by a researcher and 10 doctors who had been trained. Tests were performed on subjects in a school classroom with one-to-one interviews.

We measured internal consistency (Cronbach’s alpha and item-to-total correlation) and inter-rater reliability of the Indonesian version of SYSTEMS-R. Cronbach’s alpha was considered to be statistically significant if ≥ 0.7 .^{9,16} Item-to-total correlation was measured using point-biserial correlation (r_{pbis}). Item with item-to-total correlation ≥ 0.3 fulfilled the psychometric requirement to be included as part of the test/instrument.¹⁸ The correlation was significant for P values < 0.05 ; P value was computed based on

Table 1. Internal consistency (Cronbach's Alpha) of the Indonesian Version of SYSTEMS-R

Data I (n=133)							
Age group, years	4-5	6-7	8-9	10-11	12-13	14-15	All Ages (4-15)
Cronbach's alpha	0.836	0.818	0.752	0.263	0.639	0.737	0.936
Data II (n=133)							
Age group, years	4-5	6-7	8-9	10-11	12-13	14-15	All Ages (4-15)
Cronbach's alpha	0.871	0.817	0.818	0.428	0.705	0.780	0.941

T-test. Inter-rater reliability was determined using the Bland-Altman method.^{21,22} The measurement from two raters had good agreement if the differences were within mean \pm 1.96 SD.^{21,22} Data analysis was performed with Microsoft Excel Megastat.

The operational definition of age groupings was as follows: 4-5 year old: children aged from 4 years to 5 years and 364 days; 6-7 year old: children aged from 6 years to 7 years and 364 days; 8-9 year old: children aged from 8 years to 9 years and 364 days; 10-11 year old: children aged from 10 years to 11 years and 364 days; 12-13 year old: children aged from 12 years to 13 years and 364 days; and 14-15 year old: children aged from 14 years to 15 years and 364 days. Neurologic impairments were defined as impairments of the central nervous system manifested as seizures, epilepsy, developmental delay, or cerebral palsy. Communication impairments were defined as impairments of language manifested as expressive language delay (speech delay) or receptive language delay (unable to comprehend speech/commands/talk).

Results

The Indonesian version of SYSTEMS-R was administered to 133 children aged 4-15 years. Ninety-eight subjects were tested once and 35 subjects were tested twice, by two different raters in a counterbalanced order on the same day. Rater 1 was a researcher and rater 2 was a doctor who was trained to administer the Indonesian version of SYSTEMS-R.

For the computation of internal consistency (Cronbach's alpha and item-to-total correlation), there were two sets of data derived from 133 subjects (68 boys and 65 girls). Data I was from the 98 subjects who were tested once plus data from the 35 subjects tested by rater I. Data II was from 98 subjects who were tested once plus data from the 35 subjects tested by rater II. Data from the 35 subjects (18 boys

Table 2. Item-to-total correlation of Indonesian version of SYSTEMS-R

Data I (n=133)			Data II (n=133)		
Item	r _{pbis}	P value	Item	r _{pbis}	P value
1	0.768	0.000	1	0.777	0.000
2	0.794	0.000	2	0.801	0.000
3	0.852	0.000	3	0.834	0.000
4	0.379	0.000	4	0.498	0.000
5	0.276	0.001	5	0.266	0.002
6	0.530	0.000	6	0.547	0.000
7	0.522	0.000	7	0.564	0.000
8	0.501	0.000	8	0.591	0.000
9	0.492	0.000	9	0.538	0.000
10	0.424	0.000	10	0.432	0.000
11	0.676	0.000	11	0.656	0.000
12	0.577	0.000	12	0.609	0.000
13	0.707	0.000	13	0.712	0.000
14	0.543	0.000	14	0.544	0.000
15	0.706	0.000	15	0.722	0.000
16	0.771	0.000	16	0.773	0.000
17	0.771	0.000	17	0.777	0.000
18	0.676	0.000	18	0.664	0.000
19	0.652	0.000	19	0.655	0.000
20	0.541	0.000	20	0.575	0.000
21	0.379	0.000	21	0.451	0.000
22	0.365	0.000	22	0.326	0.000
23	0.364	0.000	23	0.374	0.000
24	0.410	0.000	24	0.431	0.000
25	0.241	0.005	25	0.279	0.001
26	0.446	0.000	26	0.461	0.000
27	0.348	0.000	27	0.384	0.000
28	0.401	0.000	28	0.403	0.000
29	0.562	0.000	29	0.531	0.000
30	0.718	0.000	30	0.720	0.000
31	0.827	0.000	31	0.833	0.000
32	0.735	0.000	32	0.725	0.000
33	0.657	0.000	33	0.643	0.000
34	0.680	0.000	34	0.708	0.000
35	0.360	0.000	35	0.348	0.000
36	0.568	0.000	36	0.572	0.000
37	0.376	0.000	37	0.364	0.000
38	0.483	0.000	38	0.497	0.000
39	0.420	0.000	39	0.447	0.000
40	0.319	0.000	40	0.341	0.000

Note: r_{pbis}: item-to-total correlation with formula of point biserial; P value was computed based on T-test

and 17 girls) who were tested twice was used for the computation of inter-rater reliability.

Table 1 shows the internal consistency results of the Indonesian version of SYSTEMS-R. For all subjects (4–15 years age group), Cronbach’s alpha was 0.936 for Data I and 0.941 for Data II. Since Cronbach’s alpha for both data sets was >0.7, the Indonesian version of SYSTEMS-R can be considered reliable. However, Cronbach’s alpha for the 10–11 years age group in Data I and II, and the 12–13 years age group in Data I were <0.7, indicating that they did not fulfill the requirement of reliability, in terms of internal consistency.

The item-to-total correlation analysis and the

correlation of each individual item with total score²¹ is shown in Table 2. The items in Data I and II had correlations (r_{pbis}) >0.3, except for items 5 and 25. The correlation of item 5 was 0.266–0.276, and that of item 25 was 0.241–0.279. However, significance test for all items in Data I and II resulted in P values <0.05. In fact, the P values of each of the items were <0.01, hence, all items of the Indonesian version of SYSTEMS-R can be considered reliable.

Figure 1 shows the Bland-Altman plot of the scores obtained by raters 1 and 2 for 35 children aged 4–15 years. The differences were within mean \pm 1.96 SD, so there was good agreement between the scores obtained by the two raters. The difference was not significant based on Pitman’s test (P=0.077).

The scores of the Indonesian version of SYSTEMS-R obtained in this study are shown in Table 3. Based on Data I and Data II, the percentage of subjects having a score at or below the age-appropriate cut-off score was 32–35%.

Table 3. Scores of the Indonesian version of SYSTEMS-R

Score	≤Cut-off score	>Cut-off score	Total
Data I, n (%)	42 (32)	91 (68)	133 (100)
Data II, n (%)	46 (35)	87 (65)	133 (100)

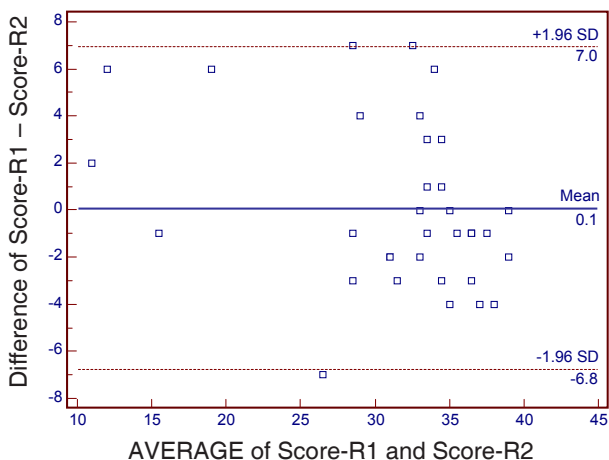


Figure 1. Bland-Altman plot of scores from the Indonesian version of SYSTEMS-R obtained by two raters

Score-R1: score obtained by rater 1

Score-R2: score obtained by rater 2

Differences:

Sample size	= 35
Arithmetic mean	= 0.0857
95% CI	= -1.1198 to 1.2912
Standard deviation	= 3.5094
Lower limit	= -6.7927
95% CI	= -8.8728 to -4.7126
Upper limit	= 6.9642
95% CI	= 4.8840 to 9.0443

Pitman’s test of difference in variance: P = 0.077 (significant if P<0.05)

Discussion

We found that the Indonesian version of SYSTEMS-R was internally consistent across age groups (4–15 years) with Cronbach’s alpha of 0.936–0.941 (n=133). However, internal consistency was not as strong for the 10–11 years age group (Cronbach’s alpha 0.263–0.428), possibly due to the exclusion of children with neurologic or communication impairments, or because some items were too easy for that age group, leading to little or no variance in the item scores, hence, a lower Cronbach’s alpha value. However, the overall Cronbach’s alpha across all age groups (4–15 years) was 0.936–0.941 (>0.7), so the Indonesian version of SYSTEMS-R can be considered reliable.

The internal consistency of the Indonesian version of SYSTEMS-R was similar to that reported by Ouvrier *et al.*^{10,15} and Bornholt *et al.*,^{9,11} who used the original SYSTEMS-R. Ouvrier *et al.*^{10,15} showed that SYSTEMS was internally consistent across age groups^{10,15} for a large sample of children from their school studies (Cronbach’s alpha coefficient =0.92; n=1,013), however, internal consistency was not as strong for groups of 10 and 11-year-old children (Cronbach’s alpha for age 5 years was 0.77, 6 years 0.83, 7 years 0.80, 8 years 0.64, 9 years 0.67, 10 years

0.52, and 11 years 0.56).¹⁵ SYSTEMS was reliable for preschool children aged 4–5 years in terms of internal consistency (Cronbach's alpha = 0.82). SYSTEMS was also reliable for children aged 5–12 years (Cronbach's alpha >0.7), so that SYSTEMS could be used for children aged 4–12 years.¹¹ Subsequently, Bornholt *et al.*^{9,12} revised the SYSTEMS in order to extend its use for children and adolescents; the revised SYSTEMS is named SYSTEMS-R. Studies in an adolescent sample (10–15 years of age) showed that the response was reliable (Cronbach's alpha >0.7) and unbiased by sex or socioeconomic indicators, therefore, SYSTEMS-R can be used for children and adolescents aged 4–15 years.^{9,12}

In the analysis of items of the Indonesian version of SYSTEMS-R, we found that all items fulfilled the psychometric requirement to be included in the test/instrument with an item-to-total correlation (r_{pbis}) >0.3, except for items 5 and 25, suggesting that the items might not be related to other items. But in the significance test, all items showed P values <0.01, thus all items in the Indonesian version of SYSTEMS-R can be considered reliable. In the selection of items to be included or eliminated from the instrument, the criteria used is not only item-to-total correlation of ≥ 0.3 , but also the content of the test domain and the aim of the test results.¹⁸ To the best of our knowledge, there have been no reports on item-to-total correlation in SYSTEMS-R.

We found good agreement between scores obtained by rater 1 and rater 2 in 35 children aged 4–15 years, based on the Bland-Altman method (the differences were within mean \pm 1.96 SD). Similarly, Ouvrier *et al.*^{10,15} reported that scores obtained by researcher 1 and researcher 2 from schoolaged children 5–11 year old (n=69) had very strong correlation ($r=0.94$).

Cognitive function screening in children has been increasingly used in many clinical and educational settings, mainly to indicate when full cognitive assessments are needed.⁹ SYSTEMS-R scores at or below age-appropriate cut-off scores would suggest cognitive impairment or cognitive deterioration and would indicate the need for a more detailed cognitive assessment.^{9,10,12,15}

The scores of the Indonesian version of SYSTEMS-R obtained in this study showed that 32–35% subjects had total scores at or below age-appropriate

cut-off scores. As such, further testing is recommended with a standardized cognitive assessment tool such as the Griffiths Mental Development Scale²³, Wechsler Intelligence Scale for Children,²⁴ or Stanford-Binet Intelligence Scale.^{23,24} Ouvrier *et al.*¹⁰ found that in 76 patients from the Neurology Clinic at Children's Hospital in Westmead, Australia, who were tested with SYSTEMS, 37 (48%) patients had scores at or below the age-appropriate cut-off score. After full cognitive assessment with the Differential Ability Scale and General Cognitive Ability, 2 of the 37 patients were found to have normal results.¹⁰

A limitation in this study was that we did not include children with neurologic or communication impairments; this may have contributed to less variance in the computation of internal consistency with Cronbach's alpha formula. Also, test-retest reliability was not measured, nor were the scores obtained compared to the standardized cognitive assessment tool, such as the Griffiths Mental Development Scale,²³ Wechsler Intelligence Scale for Children,²⁴ or Stanford-Binet Intelligence Scale.^{23,24} Hence, we do not have data on criterion-related validity, sensitivity, or specificity of the Indonesian version of SYSTEMS-R.

In conclusion, the Indonesian version of SYSTEMS-R is reliable. Further study to determine test-retest reliability and the validity of Indonesian version of SYSTEMS-R as compared to the gold standard of cognitive tests would be highly valuable.

Acknowledgments

We extend our sincere gratitude to Dr. Laurel Fisher, Dr Fiona Spencer, and Professor Robert Ouvrier as the copyright owners of SYSTEMS-R for their permission and assistance in this study; Prof. Dr. Abdurachman, MD and Prof. Dr. Herry Garna, MD, PhD for editing this article; Dr. Hadyana Sukandar, MSc for his assistance in data analysis; as well as Diba Artsiyanti E. Putri Basar, S.S, M.Si, Dian Lestari, S.Pd, and Wenny Mustikasari, S.T for their assistance in the translation process of SYSTEMS-R to the Indonesian language.

Conflict of interest

None declared.

References

1. Glascoe FP, Marks KP. Developmental-behavioral screening and surveillance. In: Kliegman RM, Stanton BF, Schor NF, St. Geme III JW, Behrman RE, editors. *Nelson textbook of pediatrics*. 19th ed. Philadelphia: Elsevier; 2011. p. 39-40.
2. Sitaresmi MN, Ismail D, Wahab A. Risk factors of developmental delay: a community-based study. *Paediatr Indones*. 2008;48:161-5.
3. Boyle CA, Decoufle P, Yeargin-Allsopp M. Prevalence and health impact of developmental disabilities in US children. *Pediatrics*. 1994;93:399-403.
4. Bailey DB, Hebbeler K, Scarborough A, Spiker D, Mallik S. First experiences with early intervention: a national perspective. *Pediatrics*. 2004;113:887-96.
5. Wiguna T, Setyawati N, Kaligis F. Uji diagnostik working memory rating scale (WMRS) versi Bahasa Indonesia dan proporsi anak sekolah dasar dengan kesulitan belajar dan defisit working memory di Jakarta. *Sari Pediatri*. 2012;14:191-7.
6. Santos LHC, Pimentel RF, Rosa LGD, Muzzolon SRB, Antoniuk SA, Bruck I. Cognitive and behavioral screening of children with learning disabilities: a preliminary study. *Rev Paul Pediatr*. 2012;30:93-9.
7. Bernal P, Estroff DB, Abouardham JF, Murphy M, Keller A, Jellinek MS. Psychosocial morbidity: the economic burden in a pediatric health maintenance organization sample. *Arch Pediatr Adolesc Med*. 2000;154:261-6.
8. Navon M, Nelson D, Pagano M, Murphy M. Use of the pediatric symptom checklist in strategies to improve preventive behavioral health care. *Psychiatr Serv*. 2001;52:800-4.
9. Bornholt LJ, Ajersch S, Fisher IH, Markham RH, Ouvrier RA. Cognitive screening for children and adolescent: general limits or ceiling effects? *J Child Neurol*. 2010;25:567-71.
10. Ouvrier RA, Hendy J, Bornholt LJ, Black FH. SYSTEMS: School-years screening test for the evaluation of mental status. *J Child Neurol*. 1999;14:772-80.
11. Bornholt LJ, Spencer FH, Fisher IH, Ouvrier RA. Cognitive screening for young children: development and diversity in learning contexts. *J Child Neurol*. 2004;19:313-7.
12. Fisher L, Spencer F, Ouvrier RA. SYSTEMS-R cognitive screening for children & adolescent. Australia: Watervale Systems; 2010. p. 1-2.
13. Spencer FH, Bornholt LJ. Investigating moderators of test-retest reliability in screening children's cognitive functioning. *Aus J Ed Dev Psych*. 2005;5:91-106.
14. Spencer FH, Bornholt LJ, Ouvrier RA. Test reliability and stability of children's cognitive functioning. *J Child Neurol*. 2003;18:5-11.
15. Ouvrier RA, Hendy J, Bornholt LJ, Black FH. SYSTEMS manual, guide for administration and scoring. Australia: the Children's Hospital at Westmead and The University of Sydney; 2000. p. 5-16.
16. Satari MH, Wirakusumah FF. Konsistensi penelitian dalam bidang kesehatan. Bandung: Refika Aditama; 2010. p. 74-5.
17. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm*. 2008;65:2276-84.
18. Azwar S. Reliabilitas dan validitas. 4th ed. Yogyakarta: Pustaka Pelajar; 2013. p. 25-136.
19. McDowell I. Measuring health: a guide to rating scale and questionnaires. 3rd ed. USA: Oxford University Press; 2006 [cited 2014 September 29]. Available from: <http://a4ebm.org/sites/default/files/Measuring%20Health.pdf>.
20. Portney LG, Watkins MP. Foundations of clinical research: application to practice. 3rd ed. New Jersey: Prentice Hall; 2008. p. 588-621.
21. Sedgwick P. Limits of agreement (Bland-Altman method). *BMJ*. 2013;346:f1630.
22. Myles PS, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. *Br J Anaesth*. 2007;99:309-11.
23. Visser L, Ruiter SAJ, van der Meulen BF, Ruijsenaars WAJJM, Timmerman ME. A review of standardized developmental assessment instruments for young children and their applicability for children with special needs. *Journal of Cognitive Education and Psychology*. 2012;11:102-7.
24. Campbell JM, Brown RT, Cavanagh SE, Vess SF, Segall MJ. *Journal of Pediatric Psychology*. 2008;33:999-1014.