Validity and Reliability of the Assessment Rubrics of High Service Skills Learning Outcome Test for Elementary School Students

Burhan Hambali*, Amung Ma'mun, Budi Susetyo, Yusuf Hidayat, Agus Gumilar
Physical Education Teacher Education. Universitas Pendidikan Indonesia, Indonesia

Article Info	Abstract
	Assessment rubrics of high service skill learning outcome test should have an adequate validity and reliability as a reliable measuring tool. This study aimed to test the validity and reliability level of the assessment rubric of high service skill learning outcome test for elementary school students. The research method used was descriptive quantitative method. The participants of this study were 50 elementary school students aged 11-12 years. The constructed assessment rubric referred to the performance test instrument involving individual performance assessment model. The analysis was administered using the content validity estimation (Lawshe's CVR), test-retest reliability, inter-rater reliability (Interclass Coefficient Correlation), and meta rubric analysis. The results of the analysis showed that the reliability and validity of the assessment instrument had shown a good criteria (p_value < 0.05) significantly. Based on the results of reliability and validity tests, it concludes that the compiled assessment rubric can be used to assess the high service skill learning outcomes in badminton learning for elementary school students.

*Correspondence Address : Jl. Dr. Setiabudhi No. 229, Bandung, West Java, Indonesia *https://ejournal.upi.edu/index.php/penjas/index*
E-mail   : burhanhambali@upi.edu

INTRODUCTION

One indicator of behavior change that occurs after the learning process isachievement of the objectives of the learning outcomes that have been determined. Learning outcomes are the level of mastery of students towards learning objectives within a certain period of time (Subarjah, 2010). Therefore, to find out the extent of student learning outcomes, it is necessary to carry out tests and measurements so that they can be measured using measuring instruments or instruments called learning outcomes tests. (Susetyo, 2011).

In the context of sports, the learning outcomes test is one of the tests that aims to measure the level of mastery of students' skills after going through the motion learning process(Hambali et al., 2021; Kumar & Kalidasan, 2013),so that helpstudents to evaluate the learning outcomes of their movement skills, as well as help teachers to measure and evaluate the effectiveness of the learning process that has been carried out (Kumar & Kalidasan, 2013).

The measuring instrument or test instrument used to measure learning outcomes should meet the eligibility criteria as a test instrument, namely (1) objective, (2) feasible, (3) reliable, and (4) valid. (Hambali, Hidayat, & Rahmat, 2020; Hambali et al., 2021; Lacy, 2011; Morrow Jr, Mood, Disch, & Kang, 2015; Yudiana, Hidayat, Hambali, Slamet, 2016). Objective means that a learning outcomes test instrument must be able to describe the actual situation, meaning that the learning outcomes test instrument can assess the ability of students objectively, then feasible, meaning that the instrument that will be used in measuring learning outcomes can collect data in accordance with the objectives to be studied. Measured, in addition, reliable and valid, meaning that the learning outcomes test instrument must have a level of stability to estimate what should be measured(Ali, 2011; Lacy, 2011; Morrow Jr et al., 2015)

Referring to this view, it is proper for a teacher to have a good and reliable learning outcome test assessment instrument, both the assigned tasks and the Assessment Rubric so that the assessment is carried out objectively and in accordance with the guidelines that have been prepared. It matches the viewKoirala, Davis, & Johnson (2008)that the assessment system consisting of tasks and rubrics is designed to measure the ability of students and teachers in pedagogical content (Koirala et al., 2008; Popham, 2011). *Task*s a question or movement task given by the teacher to determine the student's ability after completing the learning process (Popham, 2011), whereas*Rubrics*is one of the assessment instruments that teachers often use to assess student learning outcomes(Popham, 2011; Reddy & Andrade, 2010).

One of the learning outcomes test assessment instruments that are often used in the learning process of physical education and sports is a fundamental skill test instrument for playing badminton. This is evidenced by the results of research related to the test instrument for learning outcomes of playing badminton skills that have been carried out, Kumar & Kalidasan (2013) developed a skill test battery in ball badminton instrument for ages 18-23 years. Hambali et al. (2020) test the predictive validity of the test instrument for learning basic skills of playing high service badminton, precise lob, drop shot, and smash based on gender. Furthermore, Casebolt & Zhang, (2020) developed an authentic assessment instrument for playing badminton skills, and Hambali et al., (2021) developed a scoring

model on defensive lob skills by combining process and outcome assessment of strokes for elementary school students.

The global issue of the instrument for assessing badminton playing skills is an exciting issue to be studied in more depth, and it is assumed because badminton is basically a popular sport in Indonesia and has even been included in the study of the curriculum, both intra-curricular and extra-curricular in schools. From elementary school to university level.

In addition, the results of research that have been carried out have basically developed items or motion tasks to measure badminton playing skills, both in some basic skills and certain skills. While research and development related to the assessment rubric are still rarely carried out, Therefore, based on the issues and topics that have been discussed, as well as looking at the gaps in some of the current research results, the authors try to examine those related to testing the validity and reliability of the rubrics for assessing the learning outcomes of badminton playing skills, namely on high service basic skills for school students. Base.

METHOD
Design

The research was carried out by the method of quantitative description (Ali, 2011). This is assumed because, basically, the purpose of this study is to test the level of validity and reliability of the assessment rubric for high service skills learning outcomes for elementary school students.

Participants

Five expert judgments were used as participants in this study as subject meter experts(Azwar, 2012), which aims to validate the rubric that has been developed. The

criteria for the selected experts are academics and practitioners in the field of badminton games, physical education tests, and measurements, as well as physical education teachers as field practitioners. In addition, the participants used in the rubric trial were students in one of the elementary schools in the city of Bandung as many as 50 people with an age range of 11-12 years (M=11.6; SD=0.57).

Data collection technique

The assessment instrument used in this study is a type of performance assessment (Lacy, 2011; Morrow et al., 2015; Popham, 2011), with the aim of assessing students' ability to display tasks and rubrics on high service skills that have been developed in the form of observation sheets so that the rubrics that have been developed can be retested for validity and reliability. The following is the lattice of the high service assessment rubric instrument, which is presented in table 1.

Data is collected through the Planel Expert Judgment process(Azwar, 2012), which aims to assess the suitability of the assessment rubric with the tasks that have been prepared; in other words it is called the content validity process(Merkel, Mitchell, & Lee, 2016; Ozer, Fitzgerald, Sulbaran, & Garvey, 2014; Schmitt et al., 2013; Yudiana, Hidayat, Hambali, & S., 2017). In addition, the main data was collected through testing the rubric instrument for assessing high service skills that had been developed and validated by experts for elementary school students with an age range of 11-12 years.

Data analysis

All data generated in this study were analyzed using several analytical techniques, including: 1) Lawshe's Content Validity Ratio (CVR) aims to estimate the content validity of the rubric that has been developed

(Azwar, 2012; Susetyo, 2011); 2) Analysis*Interclass Correlation Coefficient* (ICC) aims to estimate interrater reliability (Lacy, 2011; Sporis, Jukic, & Vucetic, 2010; Vuleta, Sporis, Talovic, & Jeleskovic, 2010). 3) The Person Product Moment (PPM) analysis technique aims to estimate the reliability of the test-retest,

Table 1. High service skill rubrics

Dimension	Indicator	Tasks . Items	Measured taxonomy
Preparation	Body Ready Position	Stand ready to make a move	P2
		The position of the preparation of both	P2
	Kok position	How to hold a racket	P2
		How to hold	P2
	Hand Position	Racket swing preparation	P4
		Racket swing position	P4
Implemen-tation	Introduc-tory Prepara-tion	Body movement when going to hit	P4
		Unleash the satelkok to hit	P3
	Introduc-tion	Stelkok shooting with racket	P3
		Racket position when hitting	P3
Solution	Move Next	Racket position after hitting	P4
		Body movement after further	P4
	End of Move-ment	Racket position after further move-	P4
		Overall body movement	P4
The final result	Punch Direc-tional	The direction the satelkok bounces after being hit	P4

Note: P2= Basic Fundamental Movements, P3= Perceptual Ability, P4= Physical Abilities, P5= Skilled Movement. (Source: Taxonomy of the psychomotor domain, Harrow, 1972, in Morrow et al., 2015)

RESULTS AND DISCUSSION
Preparation of the Rubric for Assessment of High Service Skills Learning Outcomes.

The development of scoring criteria in the assessment rubric of the high service skill test instrument is compiled based on a current theoretical study. In this case, the assessment technique used in the rubric is a rating scale type with an interval scale of 1 to 4(Lacy, 2011; Morrow et al., 2015). Therefore, the determination of the highest score criteria will be based on the criteria of ideal movement in one item of motion task, so that in the process of determining and selecting the ideal movement in one item of motion task will be discussed through a joint FGD (focus group discussion) forum. *Subject meter expert*. This is done to see the views related to the scoring criteria in the rubric, to find one suitable and worthy criterion as the highest scoring criterion. The results of the development of the high service skill assessment rubric are presented on the appendix page.

Results of Content Validity Analysis Rubric of High Service Skills Learning Outcomes Test Assessment.

The assessment rubric that has been prepared through the theoretical review stage and the results of the FGD was validated by five subject meter experts and analyzed using the CVR analysis technique. Following are the results of the CVR analysis for the rubric for the assessment of the high service skills learning outcomes test instrument, which is presented in table 2.

Following the results of the CVR analysis in table 4.2 above, the ratio value ranges from 0.60 to 1.00. This value exceeds the minimum limit criterion value of 0.50, meaning that the ratio value is acceptable; or in other words, the description or part of the movement and the score criteria presented in Assessment rubrics can be used to assess high service or valid skills in the context of content validation.

Table 2. CVR Analysis Results of High Service Test Assessment

	No Item											
	1				2				3			
KS	4	3	2	1	4	3	2	1	4	3	2	1
M	5	5	5	5	5	5	5	5	5	5	5	5
MP	5	5	5	5	5	5	5	5	5	5	5	5
CVR	1	1	1	1	1	1	1	1	1	1	1	1
Note:	**	*	*	**	*	**	*	*	**	**	*	*

	No Item											
	4				5				6			
KS	4	3	2	1	4	3	2	1	4	3	2	1
M	5	5	5	5	5	5	5	5	5	5	5	5
MP	5	5	5	5	5	5	5	5	5	4	5	5
CVR	1	1	1	1	1	1	1	1	1	1	1	1
Note:	**	*	*	*	*	*	*	*	*	**	*	*

	No Item											
	7				8				9			
KS	4	3	2	1	4	3	2	1	4	3	2	1
M	5	5	5	5	5	5	5	5	5	5	5	5
MP	5	5	5	5	5	5	5	5	4	5	5	5
CVR	1	1	1	1	1	1	1	1	0.6	1	1	1
Note:	*	**	*	*	*	*	*	*	*	*	*	*

	No Item											
	10				11				12			
KS	4	3	2	1	4	3	2	1	4	3	2	1
M	5	5	5	5	5	5	5	5	5	5	5	5
MP	4	5	5	5	4	5	5	5	5	5	5	5
CVR	0.6	1	1	1	0.6	1	1	1	1	1	1	1
Note:	*	*	*	**	*	**	*	*	*	*	*	*

	No Item							
	13				14			
KS	4	3	2	1	4	3	2	1
M	5	5	5	5	5	5	5	5
MP	4	5	5	5	5	5	5	5
CVR	0.6	1	1	1	1	1	1	1
Note:	*	*	*	*	*	*	*	*

Note: KS = Score Criteria, M = Number of Assessors, MP = Number of Experts Who Declare Suitable, CVR = Content validity ratio, * Valid Criteria, ** Valid Criteria (Revised),

Reliability Testing Assessment rubric for high service skills learning outcomes test

Two reliability techniques used in testing the reliability estimate in this study are (1) test-retest reliability with the person product moment (PPM) analysis technique, which is to see the correlation between the first test and the second test. (Azwar, 2012; Morrow Jr et al., 2015; Popham, 2011; Susetyo, 2011). The test-retest method that will be used is the same-day test-retest method, namely the repetition of tests carried out on the same day(Thomas, Nelson, & Silverman, 2005). (2) Consistency reliability by using the inter-rater reliability technique of the ICC method. In this case, the statistical analysis technique that will be used is the analysis of variance(Goodwin, 2009; Morrow Jr et al., 2015). This is assumed because, basically, the ICC calcula-

tion will see the consistency of the assessments of the three observers, the ICC test in this research analysis will use SPSS windows 21. The test-retest and ICC test analysis results are presented in Tables 3 to 5.

Table 3 Results of Retest Test Reliability Analysis

Test	Dimension	mean	SD	PPM / COS	SEM
High Service	**Whole**	**50.24**	**3.26**	**0.94**	**0.80**
	Preparation	20.69	1.29	0.85	0.50
	Implementation	13.44	1.04	0.83	0.43
	Solution	12.48	1.24	0.87	0.45
	The final result	3.63	0.49	0.60	0.31

Table 4 Correlation Matrix Between Assessors

Test	Dimension		Appraiser 1	Appraiser 2	Appraiser 3
High Service	**Whole**	**Appraiser 1**	**1.00**		
		Appraiser 2	**0.84**	**1.00**	
		Appraiser 3	**0.66**	**0.62**	**1.00**
	Preparation	Appraiser 1	1.00		
		Appraiser 2	0.70	1.00	
		Appraiser 3	0.55	0.50	1.00
	Implementation	Appraiser 1	1.00		
		Appraiser 2	0.56	1.00	
		Appraiser 3	0.48	0.16	1.00
	Solution	Appraiser 1	1.00		
		Appraiser 2	0.73	1.00	
		Appraiser 3	0.36	0.38	1.00
	The final result	Appraiser 1	1.00		
		Appraiser 2	1.00	1.00	
		Appraiser 3	1.00	1.00	1.00

The results of the test-retest reliability test in table 3 obtained that the reliability coefficient values ranged from 0.60 to 0.94. The lowest reliability coefficient value was found in the final result dimension of 0.60, while the highest reliability coefficient of 0.94 was found in the reliability coefficient of the whole series of high service skills tests. Based on these results, it proves that the overall set of tasks and rubrics for the assessment of the high service skills test has a reliable or consistent reliability coefficient.

Table 5 Results of Inter Rater Reliability Analysis High Service Assessment Rubric

Sub Test	Dimension	mean	SD	ICC	SEM
High Service	Whole	50.18	3.23	0.87	1.16
	Preparation	20.69	1.22	0.80	0.55
	Implemen-	13.44	0.98	0.68	0.55
	Solution	12.48	1.20	0.73	0.62
	The final	3.57	0.46	1.00	0.00

Note:ICC = Intraclass Correlation Coefficient, SEM = Standard Error of Measurement.

While the results of the ICC reliability test in table 5 obtained the value of the reliability coefficient ranging from 0.68 to 0.87. The lowest reliability coefficient value is found in the implementation dimension of 0.68, while the highest reliability coefficient of 1.00 is in the final punch dimension. Then for the overall high service skill set, the reliability coefficient value is 0.87. This proves that the overall rubric for assessing high service skills has a reliable and objective reliability coefficient in assessing students' ability to display high service skills.

Discovery Discussion

This study aims to test the validity and reliability of the assessment rubric of the high service skills test instrument for elementary school students. The score criteria in the high service skill assessment rubric

are based on a review of existing theories using a rating scale type assessment technique with an interval scale of 1 to 4.

The results of the content validity analysis carried out by *a subject meter expert* in the FGD process, and the CVR values ranged from 0.60 to 1.00. The value of 0.60 in the content validity results proves that one expert judges that he does not agree with the description or part of the movement and the score criteria presented in the assessment rubric. This refers to a theoretical study that CVR is influenced by the number of assessors from experts and the number of people who state that it is important in assessing the suitability of items with indicators.(Susetyo, 2011) with a ratio index ranging from -1 CVR +1(Susetyo, 2011; Yudiana, Hidayat, Hambali, Slamet, 2016),with criteria Mp < M CVR < 0 ; Mp = M CVR = 0; Mp > M CVR > 0. The content validity results support several previous studies(Hambali et al., 2021; Merkel et al., 2016; Ozer et al., 2014; Schmitt et al., 2013; Yudiana et al., 2016). Therefore, the content validity test results show that the rubric compiled to assess basic high service skills is valid and can be used to assess learning outcomes for high service basic skills for elementary school students.

In addition, the test-retest reliability test obtained a correlation coefficient of 0.94, which proves that overall, the series of task items and assessment rubrics have reliable or consistent reliability coefficients, meaning that there is consistency between the data from the first and second assessments in the test-retest implementation process. It is worth the view Susetyo (2011), who thinks that"The reliability of the measuring device is based on the stability of the answers given by the test takers. If the participant's answer remains the same or does not change

much, the test device is reliable." In addition, test-retest reliability is used to see the stability of the measurement in relation to the instrument that has been developed. (Popham, 2011).

Then for the results of the analysis of the interrater reliability estimation using the ICC analysis technique, the reliability coefficient value is 0.87. This shows that the rubric used to assess the learning outcomes of high service skills is objective, and there are similarities in the scores given by the three raters. These results support the opinionGoodwin, (2009) that state that ICC comes from MSind - MSres/MSind. In addition, Susetyo (2011) states that "the assessment results from several observers should be the same when evaluating the same object because they use the same assessment criteria."The magnitude of the reliability coefficient obtained is almost close to the perfect correlation coefficient, namely 1.00. It shows that the assessment rubric used in assessing is the same and objective. (Hambali et al., 2021). The results of this study corroborate some of the results of previous studies (Goodwin, 2009; Hambali et al., 2021; Sporis et al., 2010; Vuleta et al., 2010) and prove that the assessment rubric has a reliable and objective reliability coefficient to assess the ability of elementary school students in displaying high service skills learning outcomes.

CONCLUSION

Based on the results of testing the validity and reliability of the assessment rubric that has been developed, it can be concluded that overall the rubric has a reliable estimate of validity and reliability so that it can be used to assess learning outcomes for high service skills for elementary school students.

REFERENCES

Ali, M. (2011). *Memahami Riset Perilaku Sosial.* Bandung: Pustaka Cendika Utama.

Azwar, S. (2012). *Penyusunan Skala Psikologi* (2nd ed.). Yogyakarta: Learning Library.

Casebolt, K., & Zhang, P. (2020). An Authentic Badminton Game Performance Assessment Rubric. *Strategies,* 33(1), 8–13. https://doi.org/10.1080/08924562.2019.1680329.

Hambali, B., Hidayat, Y., & Rahmat, A. (2020). Predictive Validity of Badminton Basic Skills Learning Outcome Instrument Test Based on Gender, 21 (Icsshpe 2019), 373–375.

Hambali, B., Hidayat, Y., Yudiana, Y., Juliantine, T., Rahmat, A., Gumilar, A., & Nugraha, R. (2021). Performance Assessment Instrument Model in Defensive Lob Learning for Elementary School Students. *TEGAR: Journal of Teaching Physical Education in Elementary School*, *4*(2), 88-95.

Koirala, H. P., Davis, M., & Johnson, P. (2008). Development of a performance assessment task and rubric to measure prospective secondary school mathematics teachers' pedagogical content knowledge and skills. *Journal of Mathematics Teacher Education*, *11*(2), 127-138.

Kumar, R., & Kalidasan, R. (2013). Construction of Skill Test Battery In Ball Badminton. *IJAIR*, *2*(2), 156-164.

Lacy, A. C., & Williams, S. M. (2018). *Measurement and evaluation in physical education and exercise science*. Routledge.

Merkel, E. C., Mitchell, S. A., & Lee, S. J. (2016). Content validity of the Lee Chronic Graft-versus-Host Disease Symptom Scale as assessed by cognitive interviews. *Biology of Blood and Marrow Transplantation*, *22*(4), 752-758. https://doi.org/10.1016/j.bbmt.2015.12.026

Morrow Jr, J. R., Mood, D., Disch, J., & Kang, M. (2015). *Measurement and evaluation in human performance, 5E*. Human kinetics.

Ozer, I., Fitzgerald, S. M., Sulbaran, E., & Garvey, D. (2014). Reliability and content validity of an English as a Foreign Language (EFL) grade-level test for Turkish primary grade students. *Procedia-Social and Behavioral Sciences*, *112*, 924-929.

Popham, J. (2011). Classroom Assessment : What Teachers Need to Know Sixth Edition (Vol. 46).

Popham, W. J. (2011). Classroom Assessment : What Teachers Need to Know Sixth Edition (Vol. 46). Allyn & Bacon, A Viacom Company, 160 Gould St., Needham Heights, MA 02194; World Wide Web: http://www. abacon. com.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & evaluation in higher education*, *35*(4), 435-448.https://doi.org/10.1080/02602930902862859

Schmitt, M. A., Schröder, C. D., Stenneberg, M. S., van Meeteren, N. L., Helders, P. J., Pollard, B., & Dixon, D. (2013). Content validity of the Dutch version of the Neck Bournemouth Questionnaire. *Manual Therapy*, *18*(5), 386-389. https://doi.org/10.1016/j.math.2013.01.004

S Sporis, G., Jukic, I., Milanovic, L., & Vucetic, V. (2010). Reliability and factorial validity of agility tests for soccer

players. *The Journal of Strength & Conditioning Research*, *24*(3), 679-686.

Subarjah, H. (2010). Hasil belajar keterampilan bermain bulutangkis studi eksperimen pada siswa diklat bulutangkis FPOK-UPI. *Cakrawala Pendidikan*, *3*(3).

Susetyo, B. (2011). Menyusun tes hasil belajar. *Bandung: CV Cakra*.

Silverman, J. S., Thomas, J. R., & Nelson, J. K. (2005). Research methods in physical activity. *United States: Human Kinetics Europe Ltd*.

Vuleta, D., Sporiš, G., Talović, M., & Jelešković, E. (2010). Reliability and factorial validity of power tests for handball players. *Sport science*, *3*(1), 42-46.

Yudiana, Y., Hidayat, Y., Hambali, B., & Slamat, S. (2017, March). Content validity estimation of assessment instrument based on volleyball information system of volleyball learning: Field research. In *IOP Conference Series: Materials Science and Engineering* (Vol. 180, No. 1, p. 012230). IOP Publishing.