TEGAR 4 (2) (2021) 88-95
Journal of Teaching Physical Education in Elementary School
Available online at :
https://ejournal.upi.edu/index.php/tegar/article/view/33832
DOI: https://doi.org/10.17509/tegar.v4i2.33832

Performance Assessment Instrument Model in Defensive Lob Learning for Elementary School Students

Burhan Hambali, Yusuf Hidayat, Yunyun Yudiana, Tite Juliantine, Alit Rahmat, Agus Gumilar, Reshandi Nugraha
Department of Sports Education, Postgraduate Program, Universitas Pendidikan Indonesia

Article Info

Article History:
Received : January 2021
Reviced : February 2021
Accepted : March 2021
Available Online : May 2021

Keywords:
badminton, performance assessment instrument

Abstract

The study was aimed at developing a performance assessment instrument model for defensive lob learning. The research method used was Research & Development method for developing an assessment instrument referring to eight stages of activity, including choosing good test criteria, analyzing the sport to be tested, selecting and looking for theoretical concept literatures, selecting test items, establishing procedures, administering expert judgements, testing the instrument, and testing the estimated level of validity and reliability. The participants involved were fifth grade elementary school students aged 11-12 years. The development of instrument referred to a performance test instrument with an individual performance assessment model. The analysis consisted of the estimation of content validity (Lawshe's CVR), test-retest reliability, and inter-rater reliability (interclass coefficient correlation). The results of the analysis showed that the reliability and validity of the assessment instrument obtained a good criteria significantly (p_value <0.05). The results of reliability and validity testing conclude that the developed performance assessment instrument model can be used to measure defensive lob skills in Badminton learning for Elementary School students.

*Correspondence Address : Jl. Dr. Setiabudhi No. 229 Bandung, Jawa Barat
E-mail : burhanhambali@upi.edu

INTRODUCTION

Studies regarding the instrument for assessing the basic skills in badminton have been widely carried out, for example, Kumar & Kalidasan (2013) who developed a badminton skill test on elite athletes and Casebolt & Zhang (2020) who developed authentic assessments to measure skills in playing badminton. Besides, Chia, J. S., Chow, J. Y., Barrett, L. A., & Burns, S. F. (2019) developed a test protocol in badminton training conditions. These three studies are the research and development studies examining badminton skill test instruments on several skill and performance techniques, during the game for elite athletes, aimed to carry out the assessment process in the learning or training process.

Conceptually, assessment is a systematic procedure that includes several activities, namely collecting, analyzing, and interpreting the data or information obtained (Kusaeri & Suprananto, 2012). In the learning process, assessment is one of a series of activities that must be implemented by teachers (Suwandi, 2011). Therefore, assessment is one of the indicators in examining the learning objective achievement (Kumar & Kalidasan, 2013). One of them is by measuring learning outcomes, especially measurements in the sport skill and motor skill contexts. Measurement of sport skills and motor skills is one of the fundamental aspects in measuring a person performance (Morrow, Jackson, Dish & Mood, 2015). The measurement enables the assessment process to be carried out objectively and relevant to the actual abilities of students.

In relation to Physical Education learning, assessment is a powerful mechanism that has the potential to change learning practices (Ní Chróinín, 2017). With a structure and focus on planning the teaching and learning process, assessment in Physical Education has a positive impact on teachers and students (Ní Chrónín & Cosgrave, 2013). Therefore, assessment becomes one of the activities that must be carried out by the teacher as part of the learning process sequence.

One of the assessment models that is often used in the Physical Education learning process is performance assessment. The assessment is an approach to measure the student status based on the way they complete certain tasks (Popham, 2011). Performance assessment is conducted by observing the student activities in doing something (Sundari, 2014), while assessment technique is used to assess student skills regarding the objectives and competencies to be achieved.

The assessment strategy focuses on providing feedback to students related to the assessment criteria and planning for the next lesson (Ní Chrónín & Cosgrave, 2013). Therefore, challenges for teachers include the amount of time required to plan and difficulties in accessing sample assessments and differentiating assessments for different grade levels and abilities.

The discussion in the previous study shows that assessment is one of the indicators that has an important role to determine the success of learning. A lot of research related to performance assessment had been widely conducted. Sari, A. N., Rosidin, U., & Abdurrahman, A. (2020) developed a performance assessment instrument to measure the problem solving skills of high school students in inquiry-based physics learning. Sagita, V. V., & Rahayu, W. P. (2019) developed performance assessment in the authentic assessment context. In addition, Nugroho, W. A., Yudha, R. P., Sundari, S., & Praja, H. N. (2021) analyzed the performance assessment instrument in Physical Education learning.

However, the results of research on the performance assessment model that become an interesting issue and the most widely applied by Physical Education teachers are game-based assessment instruments or known as Game Performance Assessment Instruments (Oslin, J. L., Mitchell, S. A., & Griffin, L. L. 1998; Memmert, D., & Harvey, S. 2008; Harvey, S., Cushion, C. J., Wegis, H. M., & Massa-Gonzalez, A. N. 2010) and Team Sport

Assessment Procedure (TSAP) (Arias-Estero & Castejón, 2014).

However, the two performance assessment models can only be used for tactical assessments in the team sport context. Therefore, based on the issues and topics that have been discussed, as well as seeing the gaps in the existing research results, the authors tried to re-compile a performance assessment instrument model based on individual performance assessment in defensive lob skills for elementary school, namely a sport assessment instrument assessing a person ability individually in individual sports such as research conducted by (Kumar & kalidasan, 2013; Chia, J. S., Chow, J. Y., Barrett, L. A., & Burns, S. F. 2019; Casebolt & Zhang, 2020).

METHOD

Design

This study used the Research & Development method for developing an assessment instrument, referring to the eight stages of development activities from Morrow, *et al* (2015), including 1) selecting good test criteria, 2) analyzing the sport to be tested, 3) selecting and searching for theoretical concept literature, 4) selecting test items, 5) establishing procedures, 6) conducting expert judgment assessment, 7) testing instrument, 8) testing the estimation of the validity and reliability levels.

Participants

50 participants aged 11-13 years (M=12.1; SD=0.85), 5th Grade students in one of the elementary schools in Bandung, were involved as samples for testing the developed instrument. In addition, five expert judgments were involved as participants in the study as the subject matter experts (Azwar, 2012) to validate the developed tasks and rubrics. The selected experts were academics and practitioners in the fields of Physical Education tests, measurements, and badminton games, and Physical Education teachers as field practitioners.

Data Collection Technique

The instrument used in this study is included in the performance test type (Aryee, S., Walumbwa, F. O., Seidu, E. Y., & Otaye, L. E, 2012, Morrow, at al. 2015 & Lacy, 2014), which is aimed to measure defensive lob technique skills in badminton learning. The assessment model used was performance assessment (Popham, 2011) aimed at assessing student abilities in performing the developed tasks and scoring rubrics in observation sheet format.

Panel Expert Judgment Data were collected through observation sheets developed by researchers passing the Panel Expert Judgment process (Azwar, 2012) as a content validity procedure (Merkel, Mitchell, & Lee, 2016; Ozer, Fitzgerald, Sulbaran, & Garvey, 2014; Schmitt et al., 2013; Yudiana, Y, Hidayat Y, Hambali B, 2017). The purpose of the panel expert judgment was to examine the suitability of the prepared tasks and rubrics with the indicators and dimensions of the variables to be measured (Susetyo, 2011 & Azwar, 2012). Furthermore, the main data were collected through an assessment instrument trial that had been developed and validated by the expert on 50 fifth grade elementary school students.

Data Analysis

The analysis techniques used in this research were 1) Lawshe's CVR (Susetyo, 2011; Azwar, 2012) aimed to estimate the validity of the test content of the badminton game Basic Competence, both the process-based and result-based, 2) Pearson Product Moment (PPM) analysis technique aimed to estimate test-retest reliability, and 3) Interclass Correlation Coefficient (ICC) analysis aimed to estimate interrater reliability (Goodwin, 2001; Sporis, Jukic, Milanovic, & Vucetic, 2010; Lacy, 2011).

RESULT

1. Theoretical Trial Result of Performance Assessment Instrument Model

Theoretical trials were carried out to see the expert reviews related to the tasks that had been developed. The suitability of the indi-

cators was then assessed. The instruments developed through the theoretical study were then reviewed by five experts called Subject Matter Experts (Azwar, 2012), including badminton experts, test and measurement experts, and Physical Education teachers as practitioners. This stage is called the content validity stage. The analytical technique used in this stage was the CVR (content validity ratio) analysis technique from Lawshe. In this case, the CVR calculation was based on the importance ratio of the developed items according to experts, the item is declared fulfilling content validity if there are similarities among raters with a ratio index above 0.50 (Susetyo, 2011). Ratio index ranges $-1 \leq CVR \leq +1$ (Susetyo, 2011; Naga, D. S, 2012; Azwar, 2012). The results of the content validity analysis using the CVR technique are presented in Table 1.

Based on the results of the CVR task analysis developed on the defensive lob assessment instrument, the ratio value ranged from 0.60 to 1.00. According to the minimum threshold criteria, the obtained ratio value exceeded the 0.50 value, meaning that the ratio value was acceptable. In other words, the compiled tasks can be used as the item to measure defensive lob skills.

2. Empirical Trial Results of Performance Assessment Model Instrument

The process of the empirical test was carried out on 50 grade 5 elementary school students. Students were assessed for their performance when carrying out the defensive lob test by three raters or observers using an assessment format that had been prepared. The process was recorded by a video camera so that the observers could recheck the movements displayed by students after completing the test. The test was carried out twice on the same day. The results of the empirical trial analysis of the performance assessment instrument model are presented in Table 2.

Table 1. Content validity ratio analysis

Dimension	Indicator	Task	MP	CVR
Preparation	Body Position Preparation	Standing, ready to do a movement	5	1.00
		Gripping Technique	4	**0.60**
	Motion to Shuttlecock Direction	Perceiving the coming shuttlecock direction	4	**0.60**
		Footstep Movement	5	1.00
	Hitting Position	Hitting Stance	4	**0.60**
	Racquet Backswing	Preparation of racket position when going to hit	4	**0.60**
Implementation	Racquet Forward Swing	Racket swing movement when going to hit	5	1.00
		Racket position when going to hit	5	1.00
	Impact	Impact of shuttlecock and racquet	5	1.00
		Body movement when hitting the shuttlecock	5	1.00
Completion	Further Motion	Body movement after hitting the shuttlecock	4	**0.60**
		Racquet position after hitting the shuttlecock	5	1.00
	Final Motion	Racquet position after further motion	5	1.00
		Body movement to the ready position	5	1.00
Final Result	Accuracy of Stroke Direction	Accuracy of shuttlecock direction after hitting	5	1.00

Note: MP = Number of rater declaring suitable, CVR = Content Validity Ratio.

Table 2. Results of Descriptive Statistic Test

Subject	N	Age		LB Test	
		M	SD	M	SD
Male	30	12.0	0.85	51.0	3.1
Female	20	12.2	0.88	50.3	2.8
M + F	50	12.1	0.85	50.8	3.0

Note: M = Mean, SD = Standard Deviation, LB = Defensive Lob.

Table 3. Descriptive Statistic Score of Test Retest Results

Subject	Mean Defensive Lob	
	Trial 1	Trial 2
Male	50.3	51.7
Female	49.9	50.8
M + F	50.1	51.4

Table 4. Descriptive Statistic Score of Rater Judgement Results

Subject	Mean Defensive Lob		
	P1	P2	P3
Male	51.1	50.9	51.1
Female	50.4	50.2	50.5
M + F	50.8	50.6	50.8

Reliability Test of Performance Assessment Model Instrument

There are two types of reliability techniques used in the empirical testing of the performance assessment instrument model, namely interclass coefficient reliability and intraclass coefficient reliability (Morrow, at al, 2015; Bresciani, Oakleaf, Duncan, & Hickmott. 2009; Thomas, Nelson, Silverman.2005). In this study, the reliability estimation testing of the interclass reliability type used the test-retest model, while the intraclass reliability type used the interrater reliability model (Morrow, at al, 2005; Bresciani, Oakleaf, Duncan, & Hickmott. 2009; Thomas, Nelson, Silverman. 2005;). The results of instrument reliability testing are presented in the Table 5.

Table 5. Test-Retest Reliability Analysis Results of Performance Assessment Model on Defensive Lob Technique

Dimension	Mean 1		SD	PPM	SEM
	Trial 1	Trial 2			
Total	**50.1**	**51.4**	**3.0**	**0.91**	**0.90**
Preparation	20.9	21.1	1.4	0.90	0.44
Implementation	13.3	13.8	1.0	0.82	0.42
Completion	13.0	13.3	1.0	0.84	0.40
Final Result	3.0	3.1	0.6	0.78	0.34

Note: PPM = Pearson Product Moment, SEM = Standard Error of Measurement

Based on the results of the test-retest reliability test of the performance assessment model on the defensive lob technique, the PPM value ranged from 0.78 to 0.91. The lowest PPM value was obtained from the final stroke indicator, while the highest value was obtained from the PPM value testing of overall defensive lob indicators. Overall, the results of the analysis had a reliability coefficient value above 0.75. It means that the developed performance assessment instrument model in the defensive lob technique is declared reliable.

Table 6. Results of Inter-Rater Reliability Analysis

Sub Test	Dimension	Mean	SD	ICC	SEM
Defensive Lob	**Total**	50.76	2.95	0.95	0.66
	Preparation	20.99	1.32	0.94	0.32
	Implementation	13.55	0.94	0.91	0.28
	Completion	13.14	0.96	0.94	0.24
	Final Result	3.08	0.58	0.98	0.08

The results of the interrater reliability test on the performance assessment model of the defensive lob technique obtained consistency coefficient values ranging from 0.91 to 0.98, the lowest reliability coefficient value was found in the implementation dimension (ICC = 0.91), while the highest coefficient value was found in the final stroke dimension (ICC = 0 ,98).

Based on the results of the reliability test, the instrument model had an inter-rater consistency reliability coefficient with the ICC technique above 0.75. It means that the performance assessment instrument model in the defensive lob technique is reliable and consistent in assessing student abilities in badminton learning.

DISCUSSION

This study aimed to develop a performance assessment instrument model in badminton learning on the defensive lob technique. The assessment instrument was developed based on four main dimensions, including the preparation stage, implementation stage, completion stage, and the final stroke stage. Based on the results of the content validity analysis at the panel expert judgment stage, the CVR values ranged

from 0.60 to 1.00. The validity value proved that the experts agreed with the task that had been developed to be used in assessing badminton skills on the defensive lob technique. It refers to the minimum threshold criteria of acceptance in content validity, which is 0.50 (Susetyo, 2011).

The value of 0.60 in the content validity results indicated that there was one expert who did not agree with the task items that had been developed, of course with logical considerations and suggestions for the task item improvement. Content validation using CVR is influenced by the number of assessors from experts and the number of people stating important in assessing the suitability of items with indicators (Susetyo, 2011) with ratio index $-1 \leq CVR \leq +1$ (Susetyo, 2011; Naga, D. S. 2012; Azwar, 2012) and criteria Mp < ½ M CVR < 0; Mp = ½M CVR = 0; Mp > ½ M CVR > 0. The result of the content validity supports some previous research (Merkel, Mitchell, & Lee, 2016; Ozer, Fitzgerald, Sulbaran, & Garvey, 2014; Schmitt et al., 2013; Yudiana, Y, Hidayat Y, Hambali B, 2017).

In the test-retest reliability test of the instrument model, all the obtained values had exceeded the minimum reliability criteria (0.75). It shows that the compiled assessment instrument model is consistent and stable to measure badminton playing skills on the defensive lob technique for elementary school students. The results of the test-retest reliability test in this study support the opinion of Susetyo (2011) who states that "the reliability of the measuring instrument is based on the stability of the answers given by the test takers. If the participant answer remains the same or does not change much, the test instrument is reliable." In addition, Popham (2011) & Kusaeri & Suprananto (2012) suggest that test-retest reliability is used to see the stability of the measurement related to the instrument that has been developed.

Meanwhile, in the interrater reliability estimation analysis results using the ICC analysis technique, the estimated reliability value was above 0.70. It showed that the three raters had the same similarities in assessing the skills performed by students, which means that the developed assessment instrument is consistent and objective in measuring and assessing the student abilities. This is relevant to the opinion and results of several previous studies (Sporis, Jukic, Milanovic & Vucetic, 2010); Markovic, Dizdar, Jukic & Cardinale, 2004); Vutela, Sporis, Talovic & Jeleskovic, 2010).

The accuracy of the ICC value as interrater reliability is suitable and relevant to the formula proposed by Goodwin (2001) that the ICC comes from $MS_{ind} - MS_{res} / MS_{ind}$. The reliability coefficient obtained in the trial was close to the perfect correlation coefficient (1.00). It shows that the assessment rubric used in assessing has been the same and objective. The results are in line with the view of Susetyo (2011) that "the assessment results of some observers should be the same when evaluating the same object, because they use the same assessment criteria."

CONCLUSION

Based on the results of the arrangement and test of the developed instrument model, it concludes that the developed performance assessment instrument model can be used to measure defensive lob skills in badminton learning for elementary school students.

REFERENCE

Arias-Estero, J., & Castejón, F. (2014). Using instruments for tactical assessment in physical education and extra-curricular sports. *European Physical Education Review*, *20*(4), 525–535. https://doi.org/10.1177/1356336X14539214

Aryee, S., Walumbwa, F. O., Seidu, E. Y., & Otaye, L. E. (2012). Impact of high-performance work systems on individual-and branch-level performance: test of a multilevel model of intermediate linkages. *Journal of applied psychology*, *97*(2), 287.

Azwar, S. (2012). *Penyusunan Skala Psikologi*. Yogyakarta: Edisi ke 2. Pustaka Pelajar.

Bresciani, Oackleaf, Kolkhorst, Nebeker, Bar-

low, Duncan & Hickmott. (2009). Examining Design and Inter-Rater Reliability of a Rubric Measuring Research Quality across Multiple Disciplines. *Practical Assessment, Research & Evaluation*. 12, Volume 14

Casebolt, K., & Zhang, P. (2020). An Authentic Badminton Game Performance Assessment Rubric. *Strategies*, *33*(1), 8-13.

Chia, J. S., Chow, J. Y., Barrett, L. A., & Burns, S. F. (2019). Reliability of a novel badminton intermittent exercise protocol. *Research quarterly for exercise and sport*, *90*(4), 487-496.

Goodwin, L.D. (2001). Interreter Agreement and Reliability. *Journal Measurment in Physical education And Exercise Science*. 5. (1). 13-34.

Harvey, S., Cushion, C. J., Wegis, H. M., & Massa-Gonzalez, A. N. (2010). Teaching games for understanding in American high-school soccer: A quantitative data analysis using the game performance assessment instrument. *Physical Education and Sport Pedagogy*, *15*(1), 29-54.

Kumar, Raj.P. & Kalidasan, R. (2013). Construction and development skill test in servis among ball badminton players. *Star International Journal Physical Education*. 1. (2). 5-8.

Kusaeri & Suprananto. (2012). *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu.

Lacy, Alan C. (2011). *Measurment & Evaluation in Physical Education and Exercise Science*. San Francisco: Pearson Education, Inc.

Memmert, D., & Harvey, S. (2008). The game performance assessment instrument (GPAI): Some concerns and solutions for further development. *Journal of Teaching in Physical Education*, *27*(2), 220-240.

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.

Naga, D. S. (2012). *Teori Sekor Pada Pengukuran Mental*. Jakarta. Nagarani Citrayasa.

Ní Chróinín, D. (2017). Assessment in physical education, a sociocultural perspective, by Peter Hay and Dawn Penney. *Assessment in Education: Principles, Policy & Practice*, *24*(1), 118–119. https://doi.org/10.1080/0969594x.2016.1140122

Ní Chrónín, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, *18*(2), 219–233. https://doi.org/10.1080/17408989.2012.666787

Nugroho, W. A., Yudha, R. P., Sundari, S., & Praja, H. N. (2021). Analisis Instrumen Asesmen Unjuk Kerja pada Pembelajaran PJOK di Sekolah Dasar Kota Cirebon. *Gelanggang Olahraga: Jurnal Pendidikan Jasmani Dan Olahraga*, *4*(2), 126-141.

Oslin, J. L., Mitchell, S. A., & Griffin, L. L. (1998). The game performance assessment instrument (GPAI): Development and preliminary validation. *Journal of teaching in physical education*, *17*(2), 231-243.

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* (Iceepsy 2013), 924–929. https://doi.org/10.1016/j.sbspro.2014.01.1250

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265-273.

Sagita, V. V., & Rahayu, W. P. (2019). Developing performance assessment-based authentic assessment of business plan subject. *Jurnal Pendidikan Ekonomi Dan Bisnis (JPEB)*, *7*(1), 31-46.

Sari, A. N., Rosidin, U., & Abdurrahman, A. (2020). Developing an Instrument of Performance Assessment to Measure Problem-Solving Skills of Senior High School Students in Physics Inquiry-Based Learning. *Scientiae Educatia: Jurnal Pendidikan Sains*, *9*(1), 1-13.

Schmitt, M. A., Schröder, C. D., Stenneberg, M. S., van Meeteren, N. L. U., Helders, P. J. M., 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

Sporis, G., Jukic, I., Vucetic, V. (2010). Reliability and Factorial Validity Of Agility Test for Soccer Players. *Journal of strength and conditioning research*. 24. (3). 679-686.

Sundari, S. (2014). Model Pengembangan Asesmen Kinerja (Performance Assessment) Mata Pelajaran Ipa Berbasis Nilai Karakter Di Smp Kota Ternate Maluku Utara. *Jurnal Edubio Tropika*, *2*(1)

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

Suwandi, S. (2011). *Model-model Asesmen dalam Pembelajaran*. Surakarta: Yuma Pustaka.

Thomas, Nelson, Silverman. (2005). *Research Methods in Physical Activity*. United. States. America: Human Kinetics.

Makovic, Dizdar, Jukic, & Cardinale. (2004). Reliability and Factorial Validity of Squat and Coutermovement Jump Tests. *Jurnal of Strength & Conditioning Association*. 18. (3), 551-555.

Vuleta, Sporis, Talovic & Jeleskovic (2010). Reliability and Factorial Validity Of Power Tests For Handball Players. *Journal Sport Science.* 3. (1). 42-46.

Yudiana, Y, Hidayat Y, Hambali B, S. S. (2017). Content Validity Estimation of Assessment Instrument Based on Volleyball Information System of Volleyball Learning: Field Research. *IOP Publishing IOP Conf. Series: Materials Science and Engineering 180*, *180*(1), 1–7. https://doi.org/10.1088/1742-6596/755/1/011001