

Validity and Reliability of Comprehensive Assessment Instruments for Handball and Badminton Games in Physical Education

Norkhalid Salimin¹, Julismah Jani¹, Mohd Izwan Shahril¹ & Gunathevan Elumalai¹

¹ Faculty Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia

Correspondence: Norkhalid Salimin, Faculty Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia. Tel: 60-15-4811-7197. E-mail: norkhalid@fsskj.upsi.edu.my

Received: June 15, 2015 Accepted: August 15, 2015 Online Published: September 17, 2015

doi:10.5539/ass.v11n23p12

URL: <http://dx.doi.org/10.5539/ass.v11n23p12>

Abstract

This study aimed to determine the validity and reliability of the Comprehensive Assessment instruments for handball and badminton games in physical education. This study conducted by six types of instruments in comprehensive assessment, that are handball cognitive assessment, handball psychomotor assessment, handball affective assessment, badminton cognitive assessment, badminton psychomotor assessment and badminton affective assessment. The measuring instruments of this study are built based on the level of Bloom's taxonomy (1956) for cognitive domain, taxonomy Dave (1970), taxonomy for psychomotor domain and taxonomy Krathwohl et al. (1964) for affective domain. The results showed the validity of a comprehensive assessment for handball was $r = .82$ and $r = .80$ for badminton. While the reliability of handball cognitive assessment $r = .78$ ($n = 36$), handball psychomotor assessment $r = .93$ ($n = 31$), handball affective assessment $r = .83$ ($n = 39$), badminton cognitive assessment $r = .75$ ($n = 40$), badminton psychomotor assessment $r = .81$ ($n = 40$), badminton affective assessment $r = .81$ ($n = 40$). The percentage of agreement between examiners (inter observer agreement) for handball is (70.11 %, $SD = 0.57$) and badminton (70.03 %, $SD = 0.68$). Based on the findings, this comprehensive assessment is suitable to be used as a standard instrument to assess students' achievement in handball and badminton games through Physical Education subject in School Based Assessment.

Keywords: physical education, assessment, badminton, handball, validity, reliability

1. Introduction

Physical Education (PE) is a core and compulsory subject to be taught in the Integrated Curriculum for Secondary School based on Education Act 1996, through the Professional Circular Number 25/1998 (Ministry of Education Malaysia, 1998). Physical Education plays a very significant role in contributing to the comprehensive growth and development of students via the learning experiences based on cognitive, psychomotor and affective domains (Darst & Pangrazi, 2006; Abdullah Sani, 2003; Freeman, 2001; Daeur & Pangrazi, 1995). In order to determine the achievement of learning, a form of measurement and evaluation should be carried out during the teaching process.

Teachers are required to conduct an assessment to determine the achievement of goals and objectives of PE subject. According to Bhasah (2007), evaluations are designed to assess the status of an evaluated object and to compare the status with respect to a set of standard or criteria for decision making. In this context, evaluation is a process that includes objective determining, information gathering, information processing and conclusion forming. When all these processes are conducted systematically and scientifically, the decision will be more accurate and will meet the purpose of evaluation results (Abu Bakar & Bhasah, 2008).

Since 1989, formal summative assessment in Integrated Curriculum for Secondary School (ICSS) had been introduced for PE subject and for the first time PE marks and grades are included in the students' report cards based on the assessment results. The assessment consists of two parts, that are examination methods and National Physical Fitness Standard (SEGAK) test based on Professional Circular No. 4/2008, Ministry of Education, Malaysia (2008). This assessment is actually not complete because students are only assessed by means of the examinations (cognitive domain) and SEGAK test (psychomotor domain) which focus on the fitness aspect only.

Therefore, PE subject evaluation implemented now is considered not complete and holistic, not balanced and

comprehensive due to the lack of standardized instruments and standards to be adopted by teachers to assess students in PE, especially concern to game skills. Teachers should use a standard assessment process for effective assessment. This study recommends an effective comprehensive assessment which includes a thorough assessment of cognitive, psychomotor and affective domains of learning.

The objectives of the study are as follows:

- 1) Identifying the content validity of the comprehensive assessment instruments for handball and badminton in Physical Education.
- 2) Identifying the reliability of the comprehensive assessment instruments for handball and badminton in Physical Education.
- 3) Identifying the reliability of inters observer agreement of comprehensive assessment instruments for handball and badminton in Physical Education.

This comprehensive assessment instrument is constructed based on the level of Bloom's taxonomy (1956) of the cognitive domain, taxonomy Dave (1970), taxonomy for psychomotor domain and taxonomy Krathwohl et al. (1964) for affective domain. Cognitive domain refers to the thinking and the intellect power in which the cognitive evaluation is to measure the level of knowledge and intelligence of the students (Kamarudin & Siti Hajar, 2004) and it happen all the time and everywhere (Abu Bakar & Bashah, 2008). There are six levels of cognitive classification based on Bloom's taxonomy (1956), namely (i) knowledge; (ii) comprehension; (iii) application; (iv) analysis; (v) synthesis and (vi) evaluation.

Psychomotor domain refers to skills related to the physical movement of a person. Physical Education teachers always focus on controlling movements during trainings and games. Psychomotor assessment domain measures the physical, motor, fitness, ability and efficiency (Jansma & French, 1994). In the process of teaching and learning, psychomotor domain is very significant. There are five levels of hierarchy based on Dave taxonomy (1970), namely (i) imitation; (ii) manipulation; (iii) precision; (iv) articulation and (v) naturalization.

Affective domain involves spiritual aspects with the emphasis on growth and development of attitudes, feelings, emotions and values. Feelings, attitudes and values are things learnt and they grow from time to time. If the environment is healthy, then feelings, attitudes and values inculcated will be positive (Abu Bakar, 1985). Krathwohl et al. (1964) classified the affective domain into five taxonomic levels, which are (i) receive; (ii) respond; (iii) value; (iv) organize and (v) characterize by value set.

Based on Table 1, the theoretical framework of this study is based on Bloom's taxonomy (1956) of cognitive domain, taxonomy Dave (1970) for psychomotor domain and taxonomy Krathwohl et al. (1964) for affective domain. The comprehensive assessment of this study is built based on these basic theory and it contains six types of assessments, that are cognitive, psychomotor and affective assessments for both handball and badminton games.

Table 1. Theoretical framework review

Bloom's Theory (1956)	Dave's Theory (1970)	Krathwohl Theory (1964)
Knowledge	Imitation	Receive
Comprehension	Manipulation	Respond
Application	Precision	Value
Analysis	Articulation	Organize
Synthesis	Naturalization	Characterize
Evaluation		
Comprehensive Assessment		

2. Methodology

2.1 Sample

The sample consists of 16 teachers, 40 Form 2 students from the selected schools in Larut, Matang and Selama, Perak. There are 10 expert panels involved in this study. Subject teachers and expert panels in this study are chosen using purposive sampling while the selection of students is conducted with intact, in which the teacher select a class to teach Form 2 PE and all of the students in the class become the subject of the study.

2.2 Instrument

This study uses six different types of instruments that are handball cognitive assessment, handball psychomotor assessment, handball affective assessment, badminton cognitive assessment, badminton psychomotor assessment and badminton affective assessment. Comprehensive assessments in this study are instruments designed by the researchers based on Morrow et al. (2005). Figure 1 shows the flow chart for the construction of comprehensive assessment instruments of Form 2 handball and badminton.

The first step in the process of building a comprehensive assessment is to review the best evaluation criteria. This study refers to the objective requirements of PE based on teaching syllabus (Curriculum development Centre, 1999), which are related to the cognitive, psychomotor and affective domains for Form 2 handball and badminton games.

The second step is the analysis of the instrument. The researchers had referred to the contents of the handball and badminton game skills based on syllabus in which there are eight basic skills for handball and six basic skills in badminton. The values to be applied and practiced in PE are the spirit of sportsmanship, fair play, tolerance, teamwork, discipline, competitiveness, leadership and participation. In this study, the aspect of fair play is evaluated in handball games while the aspect of sportsmanship is evaluated in badminton games.

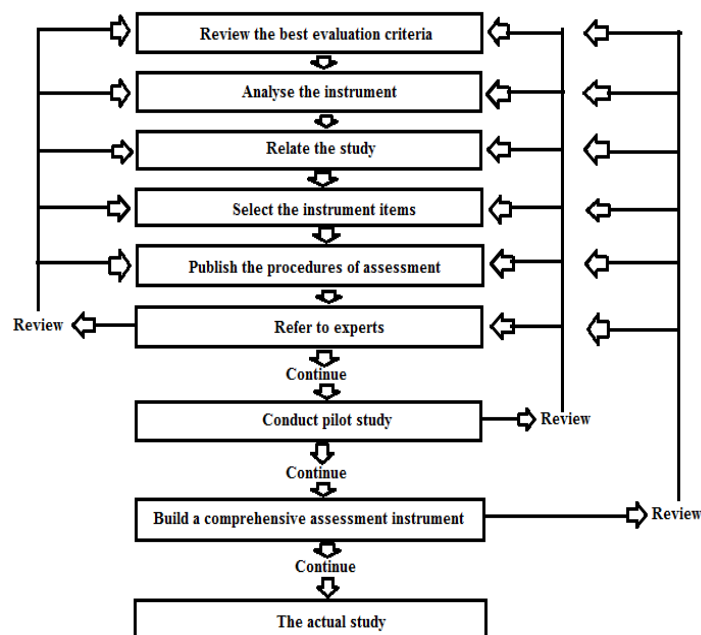


Figure 1. Flowchart of comprehensive assessment instrument construction

The third step is the study related to the construction of instrument. The comprehensive assessments of this study are based on the taxonomy of the cognitive, psychomotor and affective domains. The cognitive domain is based on Bloom's taxonomy (1956), the psychomotor domain is based on taxonomy Dave (1970) and the affective domain is based on taxonomy of Krathwohl et al. (1964).

The next step is to do the item selection for the instrument. Handball and badminton cognitive assessments consist of 40 questions and are divided into four sets of cognitive tests based on four PE teachings, involving two periods, 80 minutes per session, with each set containing 10 test questions by topic learning. This assessment is based on Bloom's taxonomy (1956), using the Table of Test Specifications and weights proposed by Hastad and Lacy (2002). Handball psychomotor assessment is divided into five basic skills including passing, catching, dribbling, checking and shooting. Badminton psychomotor assessment is divided into four basic skills; include serving, stroke, smash and footwork. Assessments are carried out by teachers during training and game sessions in the handball and badminton learning process. Handball affective assessment involves the fair game value which consists of two sub values, that are abide the rules and abide the laws. The value aspects in badminton games are the spirit of sportsmanship which consists of two sub values- accept the loss and respect the opponents.

The next step is to publish the procedures of comprehensive assessment implementation. This process is the first step in the preparation of Form 2 handball and badminton comprehensive assessment. The completed procedures are submitted to experts for review (Validity Instrument 1). At this case, the cognitive, psychomotor and affective assessments are referred to six expert panels in which two expert panels review and evaluate the content of the comprehensive assessment, two expert panels review on the aspect of language and the other two expert panels attempt and try the comprehensive assessment. A few items had been refined based on the experts' advices.

The next step is to conduct a pilot study. The pilot study was conducted on 20 Form 2 male students and 20 female students ($N = 40$) of the selected school and it involves two PE teachers. The pilot study aimed to find the validity of cognitive, psychomotor and affective assessment instruments. Some sub- assessment items had been reviewed and updated based on the feedback from the teachers during the pilot study.

The eighth step is to build a complete comprehensive assessment instruments before the actual study on comprehensive evaluation of Form 2 students in PE is conducted. At this stage, the complete comprehensive assessment is given to four expert panels to assess the validity of the content item in handball and badminton comprehensive assessment (Validity Instrument 2).

Based on the comments and recommendations of the expert panels that evaluate the content items of handball and badminton comprehensive assessment, the researcher makes some modification and item refinement to complete the comprehensive assessment. The completed comprehensive assessment is then presented to 16 PE teachers in a research workshop in the school. This one-day workshop aimed to expose the subject teachers to the procedures of the research and the assessment techniques using comprehensive assessment. Two assessment tests were carried out on the subject teacher which is the teachers' psychomotor assessment test in handball and teachers' psychomotor assessment test in badminton. The purpose of these tests is to find out the validity value of the examiners or inter observer agreement.

The final step in the process of building the comprehensive assessment instrument is to do an actual research by using the handball and badminton comprehensive assessment. The time allocated for the research on handball game is four periods of teaching sessions and is conducted in the first semester while the time allocated for the research on badminton game is four periods of teaching sessions and is conducted in the second semester.

3. Results and Discussion

3.1 The Validity of Comprehensive Assessment Instrument for Handball and Badminton

In order to find out the validity of the content of the pilot study, researchers have found some experts. The instruments' content are submitted to the experts for review. At this stage, the comprehensive assessment had been reviewed by six expert panels. Researchers have used the content validity of the questionnaire form semantic scale of 11 points with the right point marked 10 (strongly agree), the leftmost point, marked 0 (strongly disagree) and the center point is marked with 5. This questionnaire can be used to test the validity of the specific research content measuring instrument using a specialist referral (Sidek & Jamaludin, 2005). Some items have been checked and corrected based on the feedback and advices from the experts. Based on Table 2, the validity of a comprehensive assessment of the pilot study was $r = .90$ ($n = 6$). According to Abu Bakar (1985), Sidek and Jamaludin (2005), Tuckman and Waheed (1981) value of 0.70 are considered to have control or to achieve a high level. In its complete form the comprehensive assessment researchers again found some panel of experts to evaluate the comprehensive assessment content item for handball and badminton games. Four people have been referred to a panel of experts to assess the validity of a comprehensive assessment of the content item handball and badminton games.

Table 2. Content validity of the expert panel of pilot studies ($N = 6$)

Expert content	1	Expert content	2	Expert language	3	Expert language	4	Expert attempt	5	Expert attempt	6	Σ	M
9	9	8	8	9	10	8							
10	10	8	10	8	9								
9	9	8	10	9	5								
10	9	8	10	10	10								
9	10	8	9	10	10								
47	47	40	48	47	42	271	9.0						
0.94	0.94	0.80	0.96	0.94	0.84	5.42	0.90						

Based on Table 2, the result shows the validity of the item expert 1 and expert 2 for handball comprehensive assessment was $r = 0.82$ ($n = 2$), while the validity assessment of expert 3 and expert 4 for badminton comprehensive assessment was $r = 0.80$ ($n = 2$). According to Abu Bakar (1995), Sidek and Jamaludin (2005), Tuckman and Waheed (1981), value of 0.70 is considered to have mastered the highest level. There are reviews, comments and views related to the expert panel for the assessment of the handball and badminton validity assessment. With regard to all aspects of reviews, views and comments given from the expert panels, the researchers did some minor modifications to the handball and badminton comprehensive assessment so that it can truly meet the needs of the PE curriculum.

Table 3. The validity of content items of expert panels ($N = 4$)

Content Items	Validity	<i>M</i>
Handball Comprehensive Assessment	Expert 1 $r=0.84$	0.82
	Expert 2 $r=0.80$	
Badminton Comprehensive Assessment	Expert 3 $r=0.80$	0.80
	Expert 4 $r=0.80$	

3.2 The Reliability of Handball and Badminton Comprehensive Assessment Instruments

According to Ahmad (2004), Baumgartner and Jackson (1999) and Miller (2006), a reliable test gives consistent results when it is being tested repeatedly. Reliable test will produce stable and accurate data. According to Bhasah (2007), there are two procedures commonly used in estimating the reliability of test scores that are the two-test administration and one- test administration.

Cognitive assessment uses the Kuder Richardson 20 formula based on the dichotomous (true-false) scoring while the psychomotor assessment and affective assessment uses correlation method. The study was conducted on 20 male students, 20 female students ($N= 40$) and two PE teachers in the selected school. Table 3 shows the coefficient of reliability of handball and badminton cognitive assessment.

Based on Table 3, the reliability for handball cognitive test 1 is $r = 0.83$, $p = 0.65$ and $D = 0.31$ ($n = 36$), handball cognitive test 2 is $r = 0.81$, $p = 0.73$ and $D = 0.23$ ($n = 39$), handball cognitive test 3 is $r = 0.71$, $p = 0.64$ and $D = 0.38$ ($n = 35$), handball cognitive test 4 is $r = 0.76$, $p = 0.66$ and $D = 0.32$ ($n = 34$) and the mean for handball cognitive assessment is $r = 0.78$, $p = 0.67$ and $D = 0.31$. The reliability for badminton cognitive test 1 is $r = 0.75$, $p = 0.61$ and $D = 0.20$ ($n = 40$), badminton cognitive test 2 is $r = 0.70$, $p = 0.62$ and $D = 0.62$ ($n = 40$), badminton cognitive tests 3 is $r = 0.80$, $p = 0.65$ and $D = 0.65$ ($n = 40$), badminton cognitive test 4 is $r = 0.76$, $p = 0.69$ and $D = 0.24$ ($n = 40$) and the mean for badminton cognitive assessment is $r = 0.75$, $p = 0.65$ and $D = 0.43$.

Table 4. The reliability cognitive assessment for handball and badminton ($N=40$)

Instrument	Analysis Technique	<i>r</i>	<i>p</i>	<i>D</i>
Handball				
Cognitive Test 1	KR_{20}	0.83 ($n=36$)	0.65	0.31
Cognitive Test 2	KR_{20}	0.81 ($n=39$)	0.73	0.23
Cognitive Test 3	KR_{20}	0.71 ($n=35$)	0.64	0.38
Cognitive Test 4	KR_{20}	0.76 ($n=34$)	0.66	0.32
	<i>M</i>	0.78	0.67	0.31
Badminton				
Cognitive Test 1	KR_{20}	0.75	0.61	0.20
Cognitive Test 2	KR_{20}	0.70	0.62	0.62
Cognitive Test 3	KR_{20}	0.80	0.65	0.65
Cognitive Test 4	KR_{20}	0.76	0.69	0.24
	<i>M</i>	0.75	0.65	0.43

Valette (1977) suggests that the minimum instrument reliability coefficient that can be used is 0.50 while Mohd. Majid (2000) states that the reliability coefficient should be at least 0.60. Fraenkel and Wallen (1996) state that the reliability coefficient should be at least 0.70. The reliability value of handball cognitive assessment is $r = 0.78$ and $r = 0.75$ for badminton cognitive assessment, which are high enough and applicable. Difficulty index (p) is a measuring index to study the difficulty of the item being tested. Abu Bakar and Bhasah (2008), Bhasah (2007) and Morrow et al. (2005) suggest that the mean difficulty of a tested item is at 0.5, but Hastad and Lacy (2002) states that if the test is carried out to get the value of an education topic, $p > 0.50$ is acceptable. Based on Table 3, the difficulty index for handball cognitive assessment is $p = 0.64-0.73$ ($M = 0.67$) and badminton cognitive assessment is $p = 0.61-0.69$ ($M = 0.65$), which are appropriate and acceptable. Discrimination index (D) is a measurement index which describes the ability of an item to measure the sensitivity of individual differences. Ebel (1983) interprets that the discrimination index is suitable to be used to describe the role of the items in the reliability tests. Based on Table 3, the discrimination index for the handball cognitive assessment is between $D = 0.23$ to 0.38 ($M = 0.31$) and badminton cognitive assessment is between $D = 0.20 - 0.65$ ($M = 0.43$), which is appropriate and acceptable.

Table 5. The reliability psychomotor assessment for handball and badminton (N=40)

Skills	Analysis	<i>r</i>
Handball		
Overhead Pass (Training Session)	<i>Pearson product moment</i>	.79*
Bounce Pass (Training Session)	<i>Pearson product moment</i>	.92*
Jump Pass (Training Session)	<i>Pearson product moment</i>	.84*
Chest Pass (Training Session)	<i>Pearson product moment</i>	.92*
Passing (Game Session)	<i>Pearson product moment</i>	.90*
Catching (Training Session)	<i>Pearson product moment</i>	.88*
Catching (Game Session)	<i>Pearson product moment</i>	.91*
Dribble (Training Session)	<i>Pearson product moment</i>	.86*
Dribble (Game Session)	<i>Pearson product moment</i>	.92*
Checking (Training Session)	<i>Pearson product moment</i>	.91*
Checking (Game Session))	<i>Pearson product moment</i>	.95*
Jump Shot (Training Session)	<i>Pearson product moment</i>	.89*
Fall Shot (Training Session)	<i>Pearson product moment</i>	.91*
Shooting (Game Session)	<i>Pearson product moment</i>	.92*
	<i>M</i>	<i>r = .90</i>
Badminton		
Long Serve (Training Session)	<i>Pearson product moment</i>	.91*
Short Serve (Training Session)	<i>Pearson product moment</i>	.69*
Backhand Short Serve (Training Session)	<i>Pearson product moment</i>	.95*
Service (Game Session)	<i>Pearson product moment</i>	.78*
Forehand (Training Session)	<i>Pearson product moment</i>	.83*
Backhand Overhead (Training Session)	<i>Pearson product moment</i>	.65*
Forehand Overhead (Training Session)	<i>Pearson product moment</i>	.75*
Stroke (Game Session)	<i>Pearson product moment</i>	.89*
Smash (Training Session)	<i>Pearson product moment</i>	.90*
Smash (Game Session)	<i>Pearson product moment</i>	.87*
Basic Footwork (Training Session)	<i>Pearson product moment</i>	.95*
Back Footwork (Training Session)	<i>Pearson product moment</i>	.85*
Foward Footwork (Training Session)	<i>Pearson product moment</i>	.90*
Side Footwork (Training Session)	<i>Pearson product moment</i>	.79*
Foorwork (Game Session)	<i>Pearson product moment</i>	.77*
	<i>M</i>	<i>r = .83</i>

* Correlation is significant at the 0.05.

Table 6. The reliability value of handball and badminton affective assessment ($N = 40$)

Value	Analysis	r
<i>Handball</i>		
Fair play 1	<i>Pearson product moment</i>	.83*
Fair play 2	<i>Pearson product moment</i>	.82*
	<i>M</i>	$r = .83$
<i>Badminton</i>		
Sportsmanship 1	<i>Pearson product moment</i>	.83*
Sportsmanship 2	<i>Pearson product moment</i>	.78*
	<i>M</i>	$r = .81$

*Correlation is significant at the 0.05.

Table 7. The reliability inter observer agreement of handball and badminton ($N=16$)

Handball	Percentage Aggrement	<i>SD</i>	<i>Badminton</i>	Percentage Aggrement	<i>SD</i>
Slaid video 1	55.56	.68	Slaid video 1	61.10	.68
Slaid video 2	72.22	.70	Slaid video 2	61.10	.79
Slaid video 3	44.44	.73	Slaid video 3	83.30	.34
Slaid video 4	61.11	.64	Slaid video 4	55.60	.66
Slaid video 5	55.56	.59	Slaid video 5	77.80	.60
Slaid video 6	77.78	.43	Slaid video 6	83.30	.50
Slaid video 7	72.22	.54	Slaid video 7	77.80	.68
Slaid video 8	72.22	.59	Slaid video 8	50.00	.96
Slaid video 9	83.33	.38	Slaid video 9	77.80	.63
Slaid video 10	77.78	.43	Slaid video 10	77.80	.63
Slaid video 11	77.78	.47	Slaid video 11	72.20	.50
Slaid video 12	83.33	.38	Slaid video 12	72.20	.58
Slaid video 13	77.78	.69	Slaid video 13	50.00	.52
Slaid video 14	72.22	.51	Slaid video 14	44.40	.89
Slaid video 15	88.89	.32	Slaid video 15	77.80	1.15
Slaid video 16	88.89	.32	Slaid video 16	66.70	1.26
Slaid video 17	88.89	.51	Slaid video 17	61.10	.63
Slaid video 18	66.67	.55	Slaid video 18	77.80	.60
Slaid video 19	94.44	.24	Slaid video 19	72.20	.45
Slaid video 20	72.22	.46	Slaid video 20	83.30	.93
Slaid video 21	77.78	.43	Slaid video 21	72.20	.58
Slaid video 22	50.00	.71	Slaid video 22	88.90	.40
Slaid video 23	66.67	.55	Slaid video 23	61.10	.79
Slaid video 24	77.78	.47	Slaid video 24	88.90	.45
Slaid video 25	83.33	.55	Slaid video 25	77.80	.50
Slaid video 26	72.22	.59	Slaid video 26	77.80	.58
Slaid video 27	83.33	1.04	Slaid video 27	66.70	.62
Slaid video 28	66.67	.58	Slaid video 28	77.80	.70
			Slaid video 29	55.60	1.03
			Slaid video 30	61.10	.63
M	73.61	.54	M	70.37	.57

Based on Table 4, result shows that the score of 14 items in handball psychomotor assessment is between 0.79 - 0.95 ($r = 0.90$), while the score of the 15 items in badminton psychomotor assessment is between 0.65 - 0.95 ($r = 0.83$). The reliability coefficient that can be accepted by research practitioners in social science is more than $r = 0.60$. According to Mohd. Majid (2000), $r = 0.71 - 0.99$ is the best value while Fraenkel and Wallen (1996) set the acceptable reliability coefficient value at $r = 0.70 - 0.99$. Kubiszyn and Borich (2000) determine that the value of $r = 0.80 - 0.90$ is acceptable while Popham (1990) accept the value between 0.90-0.95.

Table 5 shows the values of two items in handball ($r = .83$) and badminton ($r = .81$) for the affective assessment. According to Mohd. Majid (2000), $r = 0.71-0.99$ is the best value while Fraenkel and Wallen (1996) set the acceptable value at $r = 0.70-0.99$. Kubiszyn and Borich (2000) determine that value of $r = 0.80-0.90$ is acceptable while Popham (1990), accept the value of $r = 0.90-0.95$. Based on finding, affective assessment for handball and badminton is acceptable.

In order to get the reliability of the examiners, two evaluation tests are conducted on the subject teachers ($N = 16$), that are the teachers' psychomotor evaluation in handball games and teachers' psychomotor evaluation in badminton games. These tests aimed to obtain the inter observer reliability. According to Bryington et al. (2002), there are two methods to obtain the inter observer agreement, that are the percentage agreement method and Kappa method. If the data is obtained using nominal scale, then Kappa method is suitable to be used but if there is more than one examiner for a test item, then the percentage agreement method can be used (Rink, 2002). Based on Table 6, the result shows that the percentage of agreement between the examiners (inter observer agreement) for handball assessment based on the evaluation of 28 video recordings is between 37.50 % - 93.80 % $M = 70.11$ % ($SD = 0.57$) and badminton assessment based on the evaluation of 30 video recordings is between 43.80 % - 100 % $M = 70.03$ % ($SD = 0.6$). According to Rink (2002), the acceptable reliability value of agreement between the examiners is at least 70 % (0.70).

4. Conclusion

Based on the findings, this Comprehensive Assessment is suitable to be used by teachers as a standard instrument to assess students' achievement in handball and badminton games for PE subject. The usage of this comprehensive assessment is more realistic, holistic and is able to assess students' wholly in accordance with the national education philosophy and indirectly shows the 'power of knowledge' of PE subject. This comprehensive assessment is also in line with the school-based assessment and the implementation is expected to restore the status quo of PE subject in schools throughout Malaysia.

5. Recommendations

This study involved only handball and badminton games in PE subjects in Form 2. Thus, the researcher suggests for a more comprehensive study for all levels in secondary schools. The study may involve football, ping pong and basketball games for Form 1, volleyball and softball games for Form 3 while hockey, tennis and sepak takraw games for Form 4 and basketball, cricket and rugby games for Form 5. The proposed study is intended to complement the use of Comprehensive Assessment in PE subjects for each level.

A more comprehensive study can also be done by other researchers to study the standard and level of student achievement based on learning domain of cognitive, psychomotor and affective using Comprehensive Assessment for the students of secondary schools throughout Malaysia. The aim of this study is to identify the reference norm of levels of student's learning and achievement. The standard level of achievement of learning can be identified and determined in this study. Therefore, with the scale reference norm is expected to be a reference to teachers and students in these skills. Teachers can focus on student achievement levels while students must achieve at least a minimum level in all skills learned specifically for the game in PE.

References

- Ahmad Hashim, A. (2004). *Pengukuran kecergasan motor*. Tanjong Malim: Quantum Books.
- Bakar, B. A. (2007). *Pengujian, pengukuran dan penilaian pendidikan*. Kuala Lumpur: Prospecta Printers Sdn. Bhd.
- Baumgartner, T. A., & Jackson, A. S. (1999). *Measurement for evaluation in physical education and exercise sciences* (6th ed.). USA: McGraw-Hill, Inc.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives the classification of educational goals handbook I: Cognitive domain*. New York: David McKay Company, Inc.
- Bryington, A. A., Darcy, J. P., & Marley, W. W. (2002). The estimation of interobserver agreement in behavioral

- assessment. *The Behavior Analyst Today*, 3(3), 323-328. <http://dx.doi.org/10.1037/h0099978>
- Darst, P. W., & Pangrazi, R. P. (2006). *Dynamic Physical Education for Secondary School Student* (5th ed.). Massachusetts: Allyn & Bacon.
- Dauer, V. P., & Pangrazi, R. P. (1995). *Dynamic Physical education for elementary school children* (11th ed.). Massachusetts: Allyn and Bacon.
- Dave, R. H. (1970). *Psychomotor levels developing and writing behavioral objectives*. Tuscon, AZ: Educational Innovators Press.
- Ebel, R. L. (1983). The practical validation of test of ability. *Educational measurement: Issues and Practice*, 2, 7-101. <http://dx.doi.org/10.1111/j.1745-3992.1983.tb00688.x>
- Fraenkel, J. R., & Wallen, N. E. (1996). *How to design and evaluate research in education* (3rd ed.). New York: Mc Graw Hill Inc.
- Freeman. (2001). *Physical education and sport in a changing Society* (6th ed.). San Francisco: Benjamin Cummings.
- Hastad, D. N., & Lacy, A. C. (2002). *Measurement and evaluation in physical education and exercise science* (4th ed.). Needham Height. Massachusetts: Allyn & Bacon.
- Husin, K., & Aziz, S. H. A. (2004). *Pedagogi Asas Pendidikan*. Kuala Lumpur: Kayazano Enterprise.
- Jansma, P., & French, R. (1994). *Special physical education: Physical activity, sport and recreation*. New Jersey: Prentice Hall.
- Kementerian Pelajaran Malaysia. (1988). *Surat pekeliling ikhtisas bil*. Pendidikan jasmani (Pindaan): Kuala Lumpur: Bahagian Sekolah.
- Kementerian Pelajaran Malaysia. (2008). *Surat pekeliling ikhtisas bil*. Standard kecergasan fizikal kebangsaan untuk murid sekolah malaysia (segak). Kuala Lumpur: Bahagian Sekolah.
- Konting, M. M. (2000). *Kaedah penyelidikan pendidikan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives the classification of educational goals, handbook II: Affective domain*. New York: David McKay Company, Inc.
- Kubiszyn, T., & Borich, G. (2000). *Educational testing and measurement: Classroom application and practice* (6th ed.). New York: John Wiley & Sons, Inc.
- Kurikulum, P. P. (1999). *Sukatan pelajaran rendah dan menengah*. Kuala Lumpur: Kementerian Pendidikan Malaysia.
- Miller, D. K. (2006). *Measurement by the physical educator why and how* (5th ed.). Dubuque, Iowa: Wm. C. Brown.
- Morrow, J. R., Jackson, A. W., Disch, J. G., & Mood, D. P. (2005). *Measurement and evaluation in human performance* (3rd ed.). Champaign, IL: Human Kinetics.
- Noah, S. M., & Ahmad, J. (2005). *Pembinaan modul bagaimana membina modul latihan dan modul akademik*. Penerbit UPM: Serdang, Selangor.
- Nordin, A. B. (1985). *Penilaian afektif. Kajang: Masa Enterprise*.
- Nordin, A. B., & Bakar, B. A. (2008). *Penaksiran dalam pendidikan & sains sosial*. Tanjong Malim: Penerbit UPSI.
- Popham, J. (1990). *Modern Educational Measurement. A Practitioner's Perspective* (2nd ed.). New Jersey: Prentice Hall, Englewood Cliffs.
- Rink, J. E. (2002). *Teaching physical education for learning* (4th ed.). New York: McGraw Hill.
- Tuckman, B. W., & Waheed, M. A. (1981). Evaluation an individualized science programme for community college students. *Journal of Research in Science Teaching*, 18, 489-495. <http://dx.doi.org/10.1002/tea.3660180603>
- Valette, R. M. (1977). *Modern language testing: A handbook* (2nd ed.). New York: Harcourt Brace Jovanovich.
- Yahya, A. S. (2003). *Siri pentadbiran pendidikan. Mengurus sekolah*. Pahang: PTS Publications.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).