DEVELOPING STUDENTS' SELF-ASSESSMENT AND STUDENTS' PEER-ASSESSMENT OF THE SUBJECT-MATTER COMPETENCY OF PHYSICS EDUCATION STUDENTS

[1)]*Enny Wijayanti*, [2)] Kumaidi, [3)]Mundilarto
[1)]University of Palangka Raya, [2)]University ofMuhamadiyah Surakarta,
[3)]Yogyakarta State University
[1)]ennywijayanti28@yahoo.com, [2)]kuma_426@yahoo.com, [3)]mundilarto@uny.ac.id

Abstract

This study is aimed at developing Students' Self-Assessment (SSA) and Students' Peer-Assessment (SPA) as a supplementary of the formative assessment on Physics teaching and learning, finding out the construct of SSA and SPA and finding out the effectiveness and the usefulness of SSA and SPA. The development procedure consists of two stages: development and validation. It is based on the spiral model from Cennamo and Kalk, with five phases: defining, designing, demonstrating, developing and delivering. Content validity expert judgment was measured through coefficient validity analysis from Aiken. Construct of cognitive domain was derived from learning continuum and had been validated by a focus group discussion and Delphi technique. The subjects of this study are sixth semester students of Physics Education of the University of Palangka Raya. The result of the study can be concluded as follows: 1) SSA and SPA can be used as a supplementary of the formative assessment, 2) the result of model fit, in which the GFI value is 0.996 and the SRMR value is 0.062, may be taken to indicate good fit, so developed theoretical model is supported by empirical data, 3) the group of SSA and SPA are more effective than group of non-SSA and SPA, besides students response declares SSA and SPA are reasonably effective and 4) the usefulness shows that SSA and SPA can be used as a feedback of the process and the results of the students' learning. Therefore, the results of the feedback are used to improve the learning process and the results of learning continuously.

Keywords: *students' self-assessment, students' peer-assessment, Physics competency*

Introduction

This study is focused on the discussion of classroom assessment in cognitive domain which is a part of the result of dissertation research. The classroom assessment in universities is beneficial to look at how successful students conduct learning, starting from the beginning up to the process and to give information about what ways should be conducted in order to develop less-satisfying learning outcomes. The assessment of students' learning outcomes is conducted by lecturers, so the role of lecturers is very dominant to assess the achievement of students' competency (The Law of Republic of Indonesia number 20, section 58, and verse 1, 2003 about National Education System). It is aimed at monitoring the integrated process, development and correction of learning outcomes. This is suitable for the purpose of classroom assessment which is conducted as keeping-track, checking-up, finding-out and

summing up to conclude whether the students have already mastered all set competencies (Hayat, 2010, p.5).

In accordance to the above description, the assessment which is conducted in all education levels should be based on the law of national education system. In order to implement the rule as instructed in the law and to fulfill the purpose of class assessment, formative assessment is conducted in the end of learning process. The implementation of formative assessment is aimed at knowing how deep the students have mastered a competency after following a learning process. According to Arends (2007, p.231), the purpose of classroom assessment is to diagnose knowledge, to give feedback, to evaluate and to mark. However, practically, there are still few universities giving diagnosis and feedback towards the result of the assessment.

Supporting the above explanation, Amien (1987: 99) claims that feedback must be delivered soon in the learning process in order to inform students about their work. The involvement of students in assessment is a basic of balanced assessment (Tola, 2010 p. 18). Moreover, Assessment Reform Group (1994, p. 4) states that developing learning through assessment can be done by involving the students to assess their competency themselves and to understand how to correct it. By being involved in self-assessment towards their achievement of cognitive competency, students will get information related to their difficulties in the items or attributes which are considered as hard. Thus, correction or development can be conducted to improve their learning outcomes. Attribute is a competency that must be acquired by the students to accomplish a question.

The existence of obstacles in conducting classroom assessment affects the process of giving feedback. Students cannot get the feedback soon after the learning, so they are not able to identify their difficulties. The result of assessment conducted by lecturers is used to monitor the learning process, learning development and students' achievement. Thus, another assessment is needed to supplement the formative assessment that later can overcome the obstacles so that students can get feedback about their work.

Information about the result of previous learning assessment can help improving the learning process. Based on that result, students' strengths and limitations or difficulties in following the learning process can be traced. It is in accordance to Nitko (1989, p. 447) who says that evaluation must be integrated with the learning. It means that there is a close relation between assessment and learning.

Based on the previous description, the effort to improve standard of education cannot be separated from the benefits of assessment result. The main point of the improvement is the development of learning process, so assessment system should be applied as a part of the

effort to improve the standard (Kumaidi, 2001a). This integration shows that assessment is an important component because it is inseparable from the process of education and learning. For that, students' self-assessment (SSA) is needed because it involves students in the assessment activity. SSA is an assessment approach that involves the students to honestly conduct the assessment toward their work.

Appropriate solution needs to be conducted when it deals with the emergence of obstacles in the formative assessment practices. To overcome the limitations in the formative assessment practices done by far, another assessment which involves students in assessment towards their competency achievement is again needed as a supplementary. Black and William (1998, p. 14) state that practically, peer-assessment can be used as a supplementary of previous assessment and a requirement of self-assessment. To supplement the formative assessment, students are involved in assessing their peers' competency in group after conducting it individually. This type of activity to assess each other with peers in group about competencies they have mastered is called students peer-assessment (SPA).

Nowadays, an assessment that can support classroom learning process is really needed for the assessment renewal in education. By developing SSA and SPA as a supplementary of formative assessment, it is expected to give meaningful contribution to the improvement of the process of education. Sadler, White and Frederikson (Haris, 2007, p. 28) argue that application of the SSA and SPA combination is one of the ways to improve formative assessment practices by using peer- and self-assessment. They also state that peer- and self-assessment is essential for learning.

Integrated use of SSA and SPA is rarely conducted by lecturers in a general learning process and specifically in Physics class. Thus, it is needed to review the effectiveness of SSA and SPA, whether they can improve the achievement of cognitive, affective and psychomotor competencies in Physics class. As a supplementary for formative assessment, SSA and SPA are needed to be reviewed in the part of basic construct and their development procedure.

SPA technique is adopted from Wiersma (2000, p. 3) who claims the term *group evaluation* or *peer evaluation*. Another adaptation is from Kane and Lawler (Keaten, Richardson and Elisabeth, 1993, p. 3) who purposes the term *ranking*. If it is used to detect the difficulties experienced by students, peer-rating is more appropriate because it gives detail information about the level of mastered and have not been mastered (Latham and Wexley, 1982, p. 88).

Research Method

This study applied a research and development asserted by quasi experiment. SSA and SPA integrated the measuring activity of learning outcomes with the holistic learning process. Thus, Spiral Cennamo and Kalk model (2005, p. 6) was chosen. It is often stated as five phases of development which is to: (1) define, (2) design, (3) demonstrate, (4) develop and (5) deliver.

The limited test subjects were students of Physics *Tadris* majoring *Tarbiyah* in STAIN Palangka Raya. The expanded test subjects were students of Physics Education of University of Palangka Raya who had followed the subject of Vibration of Frequency. Validation test subjects of SSA and SPA were students in the sixth semester who were following the lecture. They were divided into two classes. Class A was comparative group and class B was validation test group of SSA and SPA.

Development Process

In the development process of this study, there were define, design and demonstrate phases. The first activities were preliminary study, literature review and relevant research finding review, observation and identification of Physics learning process. Information taken from the activities was used to complete the development draft and assessment draft. This development process resulted prototype 1, prescription of SSA and SPA.

Validity Process

The draft that had been resulted in the development process was then tested for its illegibility and expert judgment, for the analysis and first revision producing prototype 2 prescription of SSA and SPA and also for the supporting instruments. The result of the analysis was then tested by limited test to look at the availability of time and instrument reliability. Limited subjects were 24 students of Physics *Tadris* majoring *Tarbiyah* of STAIN Palangka Raya. However, the analysis of content validity ratings had been done using Aiken formula through expert judgment before the instruments were used. The result showed that all instruments were contently valid. After that, analysis and second revision was done for the test result. It resulted a tentative model. Expanded test was conducted to the students of Physics Education of University of Palangka Raya who had followed the subject of Vibration of Frequency. Expanded test is used to look at composite reliability, level of difficulties and differentiator of item. The result of it was then analyzed to get the third revision.

Following the above process, instruments of SSA and SPA along with supplementary instruments were used in the validity test of SSA and SPA empirically in the learning process of Physics. Based on the empirical test, analysis using SmartPLS2.0M3 was conducted to look at construct validity and composite reliability. In order to see whether the developed model had suited the theory and application, model compatibility test was conducted using a

program named generalized structured component analysis (GesCA) (Heungsun Hwang, 2011).

Data Analysis Technique

Content validity of pre-test was conducted to look at whether all instruments used in the research were contently valid. The validity used Aiken formula (1985, p. 132-133). The measurement of rater reliability also used Aiken formula (1980, p. 957-958). Reliability of limited trial test used Cronbach Alpha. Expanded test data analysis technique included level of difficulties (LD) and discrimination index (DI). LD and DI were measured by objectives essay questions using Excel program.

Crocker and Algina (1986, p. 311), Ebel and Frisbie (1986, p. 231), Linn and Gronlund (2009, p. 356) define the difficulties of items as a proportion of the correct answers. Thus, Ebel and Frisbie (1986, p. 356) claim that the higher the difficulties index, the easier the items/questions of the test.

According to McDonald (1999, p. 78), Miller, Linn and Gronlund (2009, p. 357), Reynolds, Livingston and Willson (2010, p. 150), items discrimination or distinguishing ability of items is an index that refers to the degree or level how an item can distinguish between students who gain high score and they who get lower score in certain items/questions. Ebel and Frisbie (1986, p. 230) say if the main purpose of item selection is to optimize the test reliability, items having high discrimination should be chosen. The formula to measure the index of distinguishing ability of items is as follow.

$$\text{Discrimination Index (DI)} = \frac{\text{Upper class mean} - \text{Lower class mean}}{\text{Maximum score}}$$

The criteria used to categorize DI is adopted from Cracker and Algina (1986, p. 315). In the analysis of criteria items, if all students have already mastered indicators of the competency, the DI will be 0. However, the item is still claimed as a good item and can be used to show effectiveness of the process (Mardapi, 2012, p. 188). In this study, the data resulted from empirical validity test were used to describe the effectiveness and the application of SSA and SPA which are analyzed using Excel program. Qualitative data analysis technique in this study was used to explain the procedure of SSA and SPA development.

Discussion and Research Finding

The development of SSA and SPA in this research is focused on measuring students' cognitive competency mastery. It begins with the process of theoretical review, relevant previous research findings and observation towards Physics learning practice. The collected information was used to make the development of assessment's draft. Making a formula and

designing a draft of SSA and SPA instruments about cognitive competency was conducted through learning continuum of vibration of frequency material. The development process activity resulted prototype 1, while the validity process was generally about the test. Limited test, expanded test, analysis and revision resulted tentative model. To look at the construct, the effectiveness and use of SSA and SPA, empirical process was conducted in Physics learning.

The result of Aiken content validity test is generally summarized as seen in the following Table 1.

Tabel 1. Result Test of Content Validity Instrument

Instrument	V test	V table	Category	Result
Cognitive Scoring Guidance	0.95	0.74	valid	ALR
SSA Reflection	0.94	0.74	valid	ALR
SPA Reflection	0.90	0.74	valid	ALR

ALR: Accepted with Little Revision

Based on the Aiken content validity analysis as described above, the result showed that all instruments used in the research were contently valid. The result of descriptive analysis towards all instruments was all categorized as very good. The above table shows that all rater reliability values in the test process have attended the minimum reliability as required which have more than 0.70. It means that all instruments are reliable based on the test among the raters. Limited test is aimed at knowing the clarity of each item or question, the availability of time and instruments reliability. The value of instruments reliability measured by Alpha formula is 0.73. The result of reliability coefficient among raters is shown in Table 3.

Tabel 3. Pre-test Reliability among Raters

No	Instrument	Reliability Coefficient Value among Raters (R)	Description
1.	Scoring Guidance	0.94	Reliable
2.	SPA Reflection	0.75	Reliable
3.	SSA Reflection	0.96	Reliable

Expanded test analysis included the difficulty index, discrimination index and composite reliability were then conducted as shown in Table 4 and Table 5. The result of

instruments reliability of SSA and SPA in the second test or expanded test is shown in Table 4. In Table 5, the value of level of difficulties (LD) for all items is in the range of 0.30-0.70. Thus, all items were accepted.

Tabel 4. Second Test Composite Reliability

No.	Instrument	Composite Reliability
1.	COG	0.751
2.	SSA	0.785
3.	SPA	0.790

The third test was conducted in the Physics learning process. The test subjects to validate SSA and SPA were Class A, the control group, and class B, the test group. Empirical test was used to see the construct validity of SSA and SPA, composite reliability and achievement of cognitive competency.

Tabel 5. Level of Difficulties Index (LD) and Cognitive Item Discrimination Index (DI)

	Material	Item Characteristics			Category		Desc.
		Item	LD	DI	LD	DI	
Vibration (VB)	SAP 1 Simple Harmonic Oscillation on Spring	1	0.300	0.423	M	VH	Acc
		2	0.430	0.731	M	VH	Acc
		3	0.308	0.776	M	VH	Acc
	SAP 2 Simple Harmonic Oscillation on Mathematic Swing	4	0.430	0.769	M	VH	Acc
		5	0.498	0.657	M	VH	Acc
		6	0.576	0.444	M	VH	Acc
Wave (WV)	SAP 3 Vibration Equivalence	7	0.320	0.923	M	VH	Acc
		8	0.459	0.538	M	VH	Acc
		9	0.504	0.399	M	H	Acc
	SAP 4 Wave	10	0.530	0.548	M	VH	Acc
		11	0.515	0.462	M	VH	Acc
		12	0.420	0.757	M	VH	Acc
		13	0.384	0.657	M	VH	Acc
	SAP 5 Wave Energy	14	0.638	0.581	M	VH	Acc

		15	0.471	0.487	M	VH	Acc
		16	0.633	0.650	M	VH	Acc
		17	0.583	0.615	M	VH	Acc
Sound (SD)	SAP 6 Level of Sound Intensity	18	0.693	0.731	M	VH	Acc
		19	0.697	0.846	M	VH	Acc
		20	0.627	0.256	M	H	Acc
		21	0.698	0.333	M	H	Acc
	SAP 7 Doppler Effect	22	0.694	0.469	M	VH	Acc
		23	0.689	0.633	M	VH	Acc
		24	0.697	0.628	M	VH	Acc
		25	0.700	0.550	M	VH	Acc
Average			0.540	0.595	M	VH	

Tabel 6. Composite Reliability Value of Validity Test

No.	Construct	Component	Composite Reliability
1.	COG	VB	0.735
		WV	0.745
		SD	0.773
		COG	**0.852**
2.	SSA	VB	0.788
		WV	0.822
		SD	0.744
		SSA	**0.876**
3.	SPA	VB	0.866
		WV	0.795
		SD	0.710
		SPA	**0.873**

As seen in Table 6 above, composite reliability values of validation test of the instruments of COG, SSA and SPA are 0.852, 0.876 and 0.873. It can be concluded that all instruments are reliable. Some values of T-statistics in the analysis result of construct validity of SSA and SPA instruments, component VB, WV and SD are <1.96. However, the values

are positive. Thus, they are maintained in the instruments because deleting them can decrease the reliability level. Moreover, the instruments have been claimed as valid through expert judgment. In short, all indicators of SSA and SPA are valid. When it is viewed from the test of result for inner weights, the instruments results of SSA, SSP and COG is shown in Table 7.

Tabel 7. Result for Inner Weights

	Component	Original Sample (O)	Sample Mean (M)	Std. Deviation (STDEV)	Std. Error (STERR)	T-Statistics
SSA	SSA -> VB	0.820	0.845	0.036	0.036	22.686
	SSA -> WA	0.915	0.909	0.028	0.028	33.236
	SSA -> SD	0.796	0.809	0.046	0.046	17.134
SPA	SPA -> VB	0.871	0.876	0.032	0.032	26.814
	SPA -> FR	0.895	0.905	0.024	0.024	36.998
	SPA -> SD	0.714	0.747	0.054	0.054	13.219
COG	COG -> VB	0.836	0.849	0.036	0.036	22.983
	COG -> FR	0.850	0.871	0.028	0.028	30.430
	COG -> SD	0.868	0.892	0.032	0.032	27.533

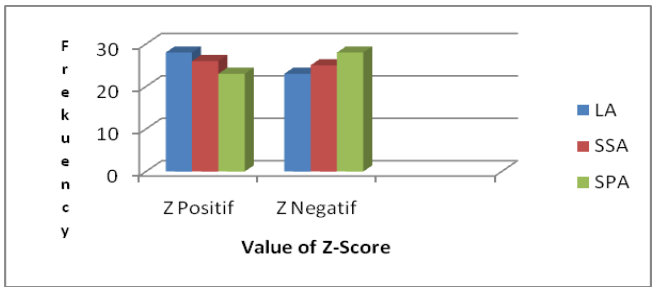
Table 7 above shows that all components of SSA, SPA and COG have T-statistics value > 1.96. Thus, generally it can be concluded that all instruments of SSA, SPA and COG are valid.

The result of fit test of SSA and SPA model shows that GFI value is 0.994 and SRMR value is 0.062. It fits well because the GFI value is closed to 1 and SRMR value ≤ 0.08. It can be concluded that the developed theoretical model is supported by empirical data. The following Table 8 shows that SSA and SPA (B) group is better than non-SSA and SPA (A) group. There is a meaningful trend of achievement increasing even though in the enough category.

Tabel 8. Achievement of Cognitive Competency

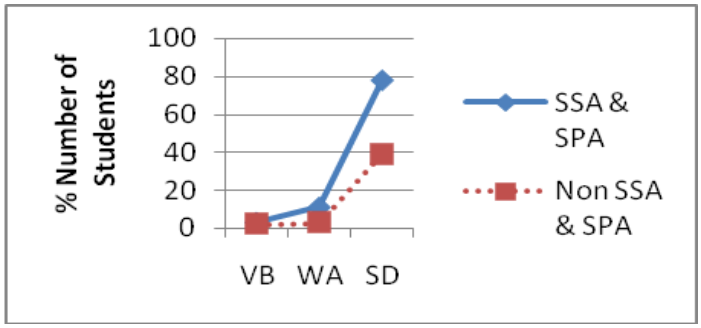
Learning Material	Max. Score	Group Average Score		Achievement Average (%)	
		A	B	A	B
1. Oscillation on Spring	30	3.4	10	11	33.3
2. Mathematic Swing	26	4.6	16	17.7	61.5
3. Equivalence of Vibration	60	3.7	16	16.2	26.7
4. Stationary Wave	40	4.1	15.6	10.3	39
5. Wave Energy	40	4.5	28.3	11.3	70.7
6. Level of Sound Intensity	30	5.5	22.5	18.3	75
7. Doppler Effect	42	6.5	31.3	15.5	74.5

If the values of z-score between lecturer's assessment (LA), SSA and SPA are compared, it is clear that the value of SSA is closer to LA than value of SPA to LA. This is because cognitive is latent; it is basically understood by the person him/herself.



Picture 1. Value Comparison between LA, SSA and SPA

SSA and SPA group show more standard achieving in subject matter. If the effectiveness result is viewed from students' response, it shows that SSA and SPA are effective to apply in Physics learning process. This is shown from the trend of increasing of students' achievement. The line graph below shows the differences of standard achieving between the two groups.



Picture 2. Line Graph of Standard Achieving Differences in Each Subject-Matter

From the point of view of self-reflection, the results of usefulness of SSA and SPA are about the students' difficulties of cognitive assessment related to attributes and the items concerned as hard (item number 2 and 23). Most students see that SSA and SPA give positive feedback towards the improvement of learning achievement. Though, they still face some difficulties on the cognitive assessment attribute, specifically on the attribute C2, P1, P2 and S6.

Conclusion

The research finding and discussion about developed SSA and SPA can be concluded as follow.

1. Empirical result shows that SSA and SPA are quite effective to use as a supplementary for formative assessment because students' learning outcomes increase. It is shown from a quite good achievement and the acquired standard achieving in the group of SSA and SPA. The profile of individual standard achieving is better in SSA and SPA group. It is concluded that SSA and SPA are quite effective to use as a supplementary of formative assessment, especially in Physics learning process.
2. Based on model fit test of SSA and SPA, it shows that the model fits because the value of GFI is 0.994 and the value of SRMR is 0.062.
3. Based on descriptive analysis and reflection of SSA and SPA, it shows that the usefulness of SSA and SPA is very beneficial to use as a supplementary of formative assessment. The information that is collected directly support it very well because it can be used as a feedback towards the development of learning achievement and the improvement as well.

References

Aiken, L.R. (1980). Content validity and reliability of single item or queationnaires. *Educational and Psychological Measurement*, 40, 955-959

..... (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142.

Assessment Reform Group. (1999). *Assessment for Learning: Beyond the black box.* University of Cambridge School of Education. Accessed on 30 Mei 2010 from http://www.assessment-reformgroup.org.uk

Black, P. et al. (2004). Working in side the black box: Assessment for learning in the classroom. *Phi Delta Kappan*. 86, 1, 8-14.

Cennamo, K. & Kalk, D. (2005). *Real world instructional design*. Australia: Thomson Learning.

Croker, L. & Algina, J. (1986). *Intoduction to classical and modern test theory*. Newyork, NY: CBS College Publishing.

Depdiknas. ((2003). *Undang-undang Republic Indonesia Nomor 20 pasal 58 ayat (1) Tahun 2003 tentang Sistem Pendidikan Nasional.*

Djemari Mardapi. (2012). *Pengukuran, penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Medika.

Ebel, R.L. & Frisbie, D.A. (1986). *Essential of educational measurement.* Englewood Cliffs, NJ: Prentice Hall, Inc.

Gronlund, N.E. & Linn, R.L. (1990). *Measurement and evaluation in teaching. (6th ed.)*. New York: Publishing Company.

Khan, Khalid S. et al. (2001). Formative self-assessment using multiple true-false question on the internet: feedback according to confidence about correct knowledge. *Medical Teacher, 23, 2, 158-163*. Birmingham: Taylor & Francis Ltd.

Kumaidi. (2001a). *Pengujian sebagai bagian peningkatan kualitas pembelajaran*. Pidato pengukuhan guru besar madya dalam ilmu evaluasi pada Fakultas Teknik Universitas Negeri Padang.