

Development of Conceptual Knowledge Test on Fluid Matter for Senior High School

Desi Prawita¹, Motlan Sirait², Sahyar³
{dsprawita@gmail.com¹}

Department of Physics Education, State University of Medan, Medan, Indonesia^{1, 2, 3}

Abstract. This study aimed to construct a multiple-choice conceptual knowledge test for high school in fluid subject matter according to standard qualification of a good test based on validity, reliability, difficulty index, discrimination index, and effectiveness distractor. The method of this study is Research and Development by using Borg and Gall step. The population of this study was all the students grade XII Senior High School in Medan, North Sumatera academic 2018/2019. The sample selection by stratified random sampling based on the public and private school, implementation of curriculum 2013, and accreditation. Data analysis technique used is descriptive qualitative and quantitative techniques. Qualitative data analysis includes material, construction and language while quantitative data analysis determined validity, reliability, difficulty index, discrimination index, and effectiveness of distractor. Step of this study consists of preliminary studies, research planning, design development, experts judgment, revision of expert judgment, a small group trials, revisions of small group trials, large group trials, revisions of large group trials. Furthermore, it obtained an instrument test of conceptual knowledge on Fluid in Senior High School that meets the standard test instruments.

Keywords: Validity, Reliability, Difficulty Index, Discrimination Index, Effectiveness of Distractor

1 Introduction

Assessment is an important component and cannot be separated from learning activities. Assessment is done to determine the extent to which students are able to understand and master the concept (Muslim, et al, 2017). Assessment is digitalization of the qualifications, expressing the observed qualifications via numbers and symbols (Kara & Celikler, 2015). Therefore, assessment of learning outcomes must be carried out consistently, systematically, and programmed.

Evaluation, on the other hand, is a decision making process relating to the assessed qualification, by comparing the results obtained from the assessment process with certain criteria (Özçelik in Kara & Celikler, 2015). One popular evaluation tool is used to find out whether learning outcomes are in line with the learning objectives of giving a test (Yunita, 2012). Often the tests are the assessment tools that are used for determination of the students' gains relating to the cognitive domain within the quantitative researches of education (Sönmez & Alacapinar, in Kara & Celikler, 2015). The test is very useful in education to analyze the causes of academic failure, evaluate learning outcomes, evaluate curriculum and as a means of motivating students to be a tool for making decisions such as whether students pass or not (Adeleke & Joshua, 2015).

To design effective instructional interventions, it is necessary to adequately assess students' conceptual knowledge. The major way to evaluate conceptual knowledge of students in physics education research is by means of a multiple-choice test (Lichtenberger et al, 2017; Kara & Celikler, 2015). Multiple-choice tests are the tests with objective grade which does not tend to differentiate from person to person (Gronlund & Lind in Kara & Celikler, 2015) and are able to be graded in a short time. The aim of instilling conceptual understanding in Physics is very difficult to establish. However, as claimed by Suarez, et al (2017), high - quality conceptual multiple-choice tests may be able to diagnose students' misconceptions and might later on lead to better understanding of the Physics concepts. This paper is intended to present the process of development test of Fluid conceptual knowledge in Senior High School.

In the study, it is aimed to analyzed the response data to provide the psychometric properties of the test. Analyzing whether the multiple choice of Fluid conceptual knowledge tests have fulfilled the validity, reliability, difficulty index, discrimination index, and effectiveness distractor qualification requirements.

2 Method

This research had done in several high schools in Medan City, North Sumatera. The population in this study were all students of class XII MIA from all high schools in Medan, on first semester of 2018/2019 academic year. Sampling is done by stratified random sampling technique, based on public or private schools, implementation of 2013 curriculum and having an A accreditation. Based on these criteria were obtained 1 Public and 4 Private High Schools, namely SMA Negeri 3 Medan, SMAS Al Azhar Medan, SMAS Al Ulum Terpadu, SMAS Amir Hamzah and SMAS Harapan Medan.

This study is research and development using Borg and Gall model. The whole process involved

simplified into three major phases: (i) Preparation stage; (ii) Development test of the Fluid Conceptual Knowledge; and (iii) Evaluation Stage.

Phase 1: Preparation Stage

In the preparation stage, reviews, observations, and literature were considered, together with the necessary inputs such as syllabus and observations. These were collected to serve as base data in the development of the conceptual test.

Phase 2: Development Stage

In the development stage, a number of Fluid item test are prepared to obtain the question design package. The development phase includes: Design pre-test products, Expert validation, Small group trial, at this stage the trial design of instrument is limited to one class from one school only which consist of 25 students, then test validity and reliability of instrument, at this stage the test results are analyzed for validity, reliability, difficulty level, discrimination index, and effectiveness distractor.

The first version resulted in an 50-item multiple choice of fluid conceptual knowledge, each question consisting of 5 options will include one correct solution and 4 distractors which categorized as Table I.

Content validity of the test is determined via the opinions of 3 science lecturers, a subject teacher, and a peer reviewer. After the test is applied to students, item analysis of the test is carried out by calculating the difficulty and discrimination of the questions of the test, validity and reliability survey is performed, inappropriate questions are excluded, KR-20 reliability coefficient is used in order to review the internal consistency between the points obtained from the test applied at the same time. A test with reliability coefficient 0,70 and above, is usually considered satisfying in terms of reliability (Fraenkel & Wallen, 2006).

Phase 3: Evaluation Stage

The next stage is that the test items were tested on a larger number of subjects by conducting large group trials. This trial was conducted on 161 students who had received the material before. Data obtained in large group trials will be analyzed quantitatively. At the end of the testing phase the test items that were declared acceptable were assembled into instrument Test of Fluid Conceptual Knowledge.

Table 1. Categorization of the 50 Fluid Conceptual Knowledge Test.

No.	Sub Matter	Bloom's Taxonomy of Cognitive Domain (Item Number)						Total
		C1	C2	C3	C4	C5	C6	
1.	Hydrostatic Pressure	1	2	3	4	43	5	6
2.	Pascal Law	6	7	8	44		35	5
3.	Archimedes Principle	9, 33	10	11	45	12, 36	42	8
4.	Surface Tension	13	14	15	46		34	5
5.	Capillarity	16	17	18, 47	19	20		6
6.	Viscosity	21	22, 48				39	4
7.	Ideal Fluid	23	24, 41	49	25			5
8.	Continuity Equation	26	28	27	40			4
9.	Bernoulli Principle	29, 37	30	50	38, 31		32	7

3 Result

The instrument used is a set of exam questions, student response, and answer key of conceptual knowledge test on fluid in senior high school consisting of 50 multiple choice test with 5 alternative answers (A, B, C, D, and E) tested to 161 students XII class of 5 Senior high schools in Medan.

Analysis data of this study consists of qualitative and quantitative analysis. In qualitative, items that have been reviewed by experts are processed using the CVR index with a scale of 1, 2, and 3, where if the conclusion TR (can be used without revision) scale 3, RK (can be used with minor revisions) scale 2, and PK (not yet used, still need consultation) scale 1. Categories of the results of the analysis of qualitative test items are presented in Table II.

In quantitative analysis, there are two steps. First, in small group trial that presented in Table III, the validity of the item is calculated using the point-biserial formula. Then consulted with a significant level of 5%, $N = 25$, r_{label} is 0.396. Based on the results of the analysis of the 50 items, it can be seen that the questions included in the valid category are 45

items (90%). Whereas the questions that are categorized as invalid are 5 items (10%). Reliability of small group trial is estimated to be 0.950. A test with reliability coefficient 0.70 and above, is usually considered satisfying in terms of reliability (Kara & Celikler, 2015). This means that the test used has good reliability.

Almost all items (sixteen exceptions) are in the difficulty range between 0.30 and 0.90, which is reasonable. Considering the distribution of item difficulties, we find 32% (16 items) in the too difficult range (0.00–0.29), 30% (15 items) in the intermediate difficult range (0.30–0.49), 28% (14 items) in the easy range (0.50–0.69), and 10% (5 items) in the too easy range (0.70–1.00). Compared to the suggestions by Doran (Lichtenberger et al, 2017), there is a small underrepresentation of items in the too easy range and a corresponding overrepresentation of items in the moderate range. Thus, the present test especially discriminates the range from very high to moderately low performance. The discrimination index was determined by calculating the difference of the mean scores between two equal-sized subgroups of the sample, one built from the 27% highest scorers, the other from the 27% lowest scorers (Lichtenberger et al, 2017).

The results of discrimination index of the test instruments in the small group found that 40% (20 items) in very well item (> 0.40), 24% (12 items) in well item (0.30-0.40), 22% (11 items) in intermediate distinctiveness(0.20-0.29), and 14% (7 items) in too weak item (≤ 0.19) (Kara & Celikler, 2015).

Second, in large group trial the validity of the item is calculated using the point-biserial formula. Then consulted with a significant level of 5%, $N = 161$, r_{tabel} is 0.154. Based on the results of the analysis of the 45 items, it can be seen that the questions included in the valid category are 36 items (80%). Whereas the questions that are categorized as invalid are 9 items (20%). Reliability of large group trial is estimated to be 0.817. This means that the test used has good reliability (Kara & Celikler, 2015).

Table 2. Results Categories Analysis of Test Items Qualitatively.

No.	Category	Item Number	Total Item
1.	Accepted	1, 6, 7, 10, 12, 16, 17, 20, 21, 23, 25, 26, 27, 28, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 45, 50	28
2.	Revised	2, 3, 4, 5, 11, 13, 14, 18, 19, 22, 24, 29, 31, 43, 44, 46, 47, 48, 49	19
3.	Rejected	8, 9, 15,	3

Table 3. Empirical Analysis of Small Group Trial.

No.	Validity $r_{tabel} = 0.396$	Difficulty 0.3 – 0.9	Discrimination Index ≥ 0.30	Choice (N = 25)					
				A	B	C	D	E	O
1	0.433	0.56	0.596	14	4	2	2	3	0
2	0.414	0.72	0.423	3	2	18*	2	0	0
3	0.567	0.68	0.506	2	2	17	2	2	0
4	0.429	0.80	0.256*	0	1	20*	2	2	0
5	0.577	0.56	0.596	3	4	14	2	2	0
6	0.573	0.72	0.583	2	18	2	2	1	0
7	0.548	0.72	0.423	2	3	2	18	0	0

No.	Validity	Difficulty 0.3 – 0.9	Discrimination Index ≥0.30	Choice (N = 25)					
	$r_{tabel} = 0.396$			A	B	C	D	E	O
8	-0.129*	0.28*	0.058*	10	7	1	0	7*	0
9	0.153*	0.12*	0.071*	0	10	12	0	3*	0
10	0.410	0.76	0.519	13	3	2	2	2	3
11	0.608	0.68	0.667	2	17	2	2	2	0
12	0.452	0.52	0.519	13	5	3	2	2	0
13	0.493	0.60	0.353	4	2	15	2	2	0
14	0.701	0.52	0.679	13	2	6	2	2	0
15	0.321*	0.16*	0.147*	12	0	6	4*	0	0
16	0.467	0.52	0.519	13	6	2	2	2	0
17	0.629	0.16*	0.308	10	5	4	2	2	2
18	0.516	0.28*	0.378	6	5	2	7	3	2
19	0.465	0.16*	0.308	10	4	7	2	2	0
20	0.528	0.40	0.449	4	10	7	2	2	0
21	0.441	0.64	0.429	16	3	2	2	2	0
22	0.583	0.28*	0.538	3	5	6	7	2	2
23	0.505	0.40	0.288*	7	4	10	2	2	0
24	0.519	0.48	0.442	2	12	3	6	2	0
25	0.507	0.36	0.212*	2	4	3	7	9	0
26	0.429	0.36	0.212*	2	3	6	9	3	2
27	0.507	0.36	0.372	7	5	2	2	9	0
28	0.477	0.44	0.205*	3	7	2	2	11	0
29	0.457	0.32	0.295*	8	2	7	3	5	0
30	0.401	0.44	0.365	11	3	7	2	2	0
31	0.465	0.16*	0.147*	2	4	5	6	5	3
32	0.473	0.32	0.455	2	8	4	3	5	3
33	0.586	0.24*	0.301	5	5	6	4	2	3
34	0.469	0.44	0.365	11	6	3	3	2	0
35	0.449	0.28*	0.218*	4	7	5	4	2	3
36	0.489	0.32	0.295*	4	6	3	2	8	2
37	0.293*	0.08*	0.154*	4	9	5	2*	0	5
38	0.542	0.24*	0.301	6	2	5	5	5	2
39	0.475	0.16*	0.147*	5	4	3	6	4	3
40	0.141*	0.08*	-0.006*	0	1	2*	9	11	2
41	0.399	0.28*	0.218*	7	6	5	3	2	2
42	0.480	0.64	0.429	2	2	3	16	2	0
43	0.600	0.28*	0.218*	5	3	6	2	7	2
44	0.513	0.32	0.295*	8	3	6	4	2	2
45	0.550	0.52	0.359	13	5	3	2	2	0
46	0.482	0.52	0.359	5	13	3	2	2	0
47	0.489	0.48	0.442	12	5	3	3	2	0
48	0.501	0.60	0.353	15	3	3	2	2	0
49	0.421	0.40	0.449	4	6	10	3	2	0
50	0.433	0.56	0.436	14	4	3	2	2	0

Almost all items (nine exceptions) are in the difficulty range between 0.30 and 0.90, which is reasonable. Considering the distribution of item difficulties, we find 20% (9 items) in the too difficult range (0.00–0.29), 22% (10 items) in the intermediate difficult range (0.30–0.49), 24% (11 items) in the easy range (0.50–0.69), and 33% (15 items) in the too easy range (0.70–1.00). Compared to the suggestions by Doran (Lichtenberger et al, 2017), there is a

small underrepresentation of items in the too easy range and a corresponding overrepresentation of items in the moderate range. Thus, the present test especially discriminates the range from very high to moderately low performance.

Table 4. Empirical Analysis of Large Group Trial.

No.	Validity	Difficulty	Discriminatio n Index	Choice (N = 161)					
	$r_{label} =$ 0,154	0.3 – 0.9	≥ 0.30	A	B	C	D	E	O
1	0.510	0.70	0.765	113	12	8	9	19	0
2	0.546	0.71	0.824	20	10	115	8	8	0
3	0.456	0.73	0.765	12	8	117	15	9	0
4	0.363	0.71	0.588	10	8	114	12	14	3
5	0.409	0.52	0.941	42	11	83	9	15	1
6	0.373	0.50	0.529	22	81	39	11	8	0
7	0.351	0.71	0.647	11	18	10	114	8	0
8	0.450	0.70	0.647	113	17	8	8	15	0
9	0.641	0.41	0.882	8	66	47	11	29	0
10	0.184	0.32	0.353	52	37	9	26	22	15
11	0.174	0.34	0.353	39	29	54	8	21	10
12	0.345	0.32	0.471	51	20	9	40	27	14
13	0.188	0.52	0.412	83	16	12	30	9	11
14	0.546	0.71	0.824	18	8	115	9	11	0
15	0.523	0.32	0.588	36	20	30	51	9	15
16	0.280	0.35	0.471	37	57	8	23	21	15
17	0.126*	0.11*	0.118*	37	17*	7	21	5	74
18	0.144*	0.07*	0.000*	11*	5	38	7	21	79
19	0.291	0.30	0.529	38	21	31	49	8	14
20	0.213	0.31	0.235*	8	47	50	17	29	10
21	0.127*	0.12*	0.294*	7	20*	17	30	5	82
22	0.134*	0.08*	0.059*	52	5	33	6	13*	52
23	0.145*	0.11*	0.235*	4	23	7	18*	38	71
24	0.216	0.20	0.353	36	39	21	18	33	14
25	0.450	0.70	0.647	13	12	10	11	113	2
26	0.546	0.71	0.824	115	8	8	8	8	14
27	0.335	0.51	0.412	82	11	8	23	22	15
28	0.264	0.50	0.647	11	81	17	30	9	13
29	0.107*	0.15*	0.176*	66	24*	7	42	7	15
30	0.338	0.70	0.588	10	8	113	13	17	0
31	0.058*	0.12*	0.000*	20*	5	26	7	34	69
32	0.320	0.16	0.529	37	26	29	18	40	15
33	0.111*	0.20*	0.235*	5	23	4	40	20*	69
34	0.154	0.32	0.353	52	32	8	33	21	15
35	0.464	0.53	0.882	13	86	8	10	34	10
36	0.295	0.35	0.529	57	36	21	14	19	14
37	0.409	0.30	0.529	27	30	10	49	20	15
38	0.173	0.29	0.706	27	26	11	35	47	15
39	0.318	0.30	0.588	48	41	10	33	11	15
40	0.536	0.54	0.824	87	20	23	8	11	12
41	0.071*	0.06*	0.059*	51	10*	7	29	32	32

No.	Validity	Difficulty	Discrimination Index	Choice (N = 161)					
	$r_{tabel} = 0,154$	0.3 – 0.9	≥ 0.30	A	B	C	D	E	O
42	0.620	0.50	1.000	81	13	21	23	8	15
43	0.418	0.31	0.529	50	29	8	20	39	15
44	0.401	0.52	0.235*	9	31	83	10	15	13
45	0.419	0.51	0.588	82	8	9	27	20	15

Table 5. Overall Test Result from the final 45-item-version of the test, taken from sample of 161 students.

Mean Score	Standard Error	Mean Difficulty	Reliability (KR-20)	Mean Validity	Mean Discrimination
18,11	0.54	0.402	0.82	0.32	0.38

The results of discrimination index of the test instruments in the small group found that 67% (30 items) in very well item (> 0.40), 9% (4 items) in well item (0.30-0.40), 11% (5 items) in intermediate distinctiveness (0.20-0.29), and 13% (6 items) in too weak item (≤ 0.19) (Kara & Celikler, 2015).

4 Discussion

After obtaining the results of validity, reliability, difficulty level, discrimination index, and effectiveness of distractors, then from each item was decided which were accepted, revised or rejected item.

The results of this study can be obtained that 34 items are accepted (76%), 2 items need to be revised (4%) and 9 items must be rejected (20%). The results of the study stated that the Student Conceptual Knowledge Test on Fluid Material in High School was good. Test reliability is calculated using KR-20, it is known that the reliability estimated to 0.817 means that the test used has good reliability (Kara & Celikler, 2015).

For items that have a good difficulty level is at a moderate level. From the data analysis, 20% (9 items) in the too difficult range (0.00–0.29), 22% (10 items) in the intermediate difficult range (0.30–0.49), 24% (11 items) in the easy range (0.50–0.69), and 33% (15 items) in the too easy range (0.70–1.00). Thus, the difficulty level of the Conceptual Knowledge Test on fluid material describes of high, medium, and low ranges (Lichtenberger et al, 2017).

From the results above, it was found that 67% (30 items) in very well item (> 0.40), 9% (4 items) in well item (0.30-0.40), 11% (5 items) in intermediate distinctiveness (0.20-0.29), and 13% (6 items) in too weak item (≤ 0.19). Thus, for the Conceptual Knowledge Test on fluid material test items in too weak discrimination index are 17, 18, 22, 29, 31, and 41. Test items generally have a good discriminating index if they have a point - biserial correlation coefficient ≥ 0.20 (Li & Singh, 2016).

Test participants in this large group were 161 students. So if it is calculated 5% of 161 students is 8.05, which means that in this test a distractor is effective if a minimum of 8 students are chosen. A good distractor if chosen a lot by the lower group. If

the distractors need to be revised. The results of the analysis of the effectiveness of the distractor's recapitulation showed that 9 (20%) questions were not functioning optimally and 36 (80%) questions were functioning optimally (Lichtenberger et al, 2017).

Based on the data above it is known that 36 items can be received and stored in the bank of conceptual knowledge on fluid material test in high school because it has fulfilled the validity, difficulty level, discrimination index, effectiveness distractor. There are 2 items, 20 and 44 that still need to be revised because they have not met well discrimination index. While 9 items, 17, 18, 21, 22, 23, 29, 31, 33, and 41 were rejected and could not be used because they did not meet any criteria for validity, difficulty level, discrimination index, and good effectiveness distractor (Barniol & Zavala, 2014).

Revising item is needed to improve validity, difficulty level, discrimination index, and effectiveness of distractor so all items become good value and quality items. A good quality item will be able to carry out its function as a tool for good evaluation. The application of a fine system also needs to be done to minimize the possibility of students guessing answers to multiple choice items. Giving penalties by reducing the value for each wrong answer, students will be more careful in choosing answers. If students really do not know, then students will prefer not to answer rather than just guessing answers because they fear the value will decrease (Lichtenberger et al, 2017).

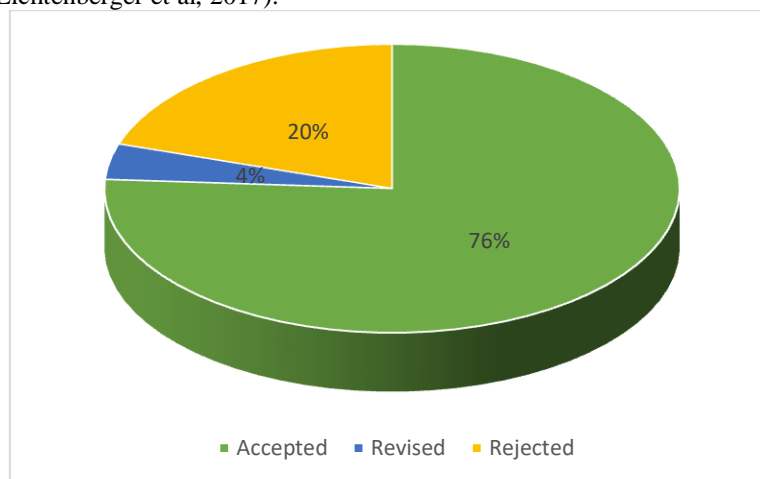


Fig. 1. Pie Diagram of Empirical Analysis in Large Group Trial.

5 Conclusion

One of the fundamental elements of a successful education is a successful assessment process. In order to carry out a successful assessment, a test with validity and reliability are ensured is required to be used. For this reason, it is aimed to develop an achievement test for Fluid Conceptual Knowledge.

Regarding to the content validity of the test, it is common in the literature that the opinions of the domain experts and lecturers to be consulted. Validity of the test which is prepared within the frame of the survey, is ensured in line with the opinions of 3 physics lecturers, 1 physics teacher, and 1 peer reviewer. As a result of the review by the domain

experts, it is determined that the content validity of the test has been provided, and is suitable for the purpose and student level that the items of which the discrimination is 0,19 and below should not be used or be reformulated, whereas that the items of which the discrimination is between 0,20-0,29 can be used as is in unavoidable circumstances, or should be corrected. Therefore, 11 questions of which the distinctiveness are below 0,30 are excluded from the test, as a result of the item analysis of the test. The final state of the test consists of 36 questions in total, 9 for each gain in the program. After the questions are excluded from the test, it is determined that the average item difficulty of the questions is intermediate (0.54) and the discrimination is in a well state (0.38). It is established that the KR-20 coefficient of the test prepared is sufficient (0.82) for the reliability of a test. As a result of the survey, a valid and reliable, multiple-choice test consisting of 45 questions of which the difficulty and discrimination is at demanded level for Fluid Conceptual Knowledge Test.

The developed test is suggested to be used by the physics teacher in order to determine the forwardness of the XII th grade senior high school students to fluid conceptual knowledge, their achievement during the training process and their misconception. In addition, the test can be used by researchers those who carry out works regarding to the effect of a certain method to the achievements relating to this knowledge.

6 Suggestion

This test instrument that has been tested and analyzed can be used as a question bank. This research shown that to make a good psychometric of the test it is necessary to endure the stages in the development of test so this will foster the motivation of the actors in the field of education to make a well physics test instrument on other material. Other researchers who want to do similar research wherever possible should take more trial subjects so as to produce more varied data. Further development can use modern test theory so that research data does not depend on the condition of students.

References

- [1] Adeleke, A. A., & Joshua, E. O.: Development and Validation of Scientific Literacy Achievement Test to Assess Senior Secondary School Students' Literacy Acquisition in Physics. *Journal of Education and Practice*. Vol 6(7), pp 28-42. (2015)
- [2] Barniol, P., & Zavala, G.: Test Of Understanding Of Vectors: A Reliable Multiple-Choice Vector Concept Test. Vol. 10(1), 010121.. *Physical Review Special Topics-Physics Education Research*. (2014)
- [3] Fraenkel, J. R. & Wallen, N. E.: *How to Design Evaluate Research in Education*. New York: McGraw-Hill Companies, 7th Edition. (2006)
- [4] Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C.: Development And Application Of A Four-Tier Test To Assess Pre-Service Physics Teachers' Misconceptions About Geometrical Optics. *Research in Science & Technological Education*. Vol. 35(2), 238-260. (2017)
- [5] Kara, F., & Celikler, D.: Development of Achievement Test: Validity and Reliability Study for Achievement Test on Matter Changing. *Journal of Education and Practice*. Vol 6(24), 21-26. (2015)
- [6] Li, J., & Singh, C.: Developing And Validating A Conceptual Survey To Assess Introductory Physics Students' Understanding Of Magnetism. *European Journal of Physics*, Vol. 38(2), 025702. (2016)
- [7] Lichtenberger, A., et al.: Validation And Structural Analysis Of The Kinematics Concept Test. *Physical Review Physics Education Research*, Vol. 13(1), 010115. (2017)

- [8] Motaz, A.: Measurement and assessment in teaching (8th ed.). Up River, NJ; Merril. (2000)
- [9] Loverude, M. E., Kautz, C. H., & Heron, P. R. Helping Students Develop An Understanding Of Archimedes' Principle. I. Research on student understanding. American Journal of Physics, 71(11), 1178-1187. (2011)
- [10] Motaz, A.: Start programming using Object Pascal. Vol. 2, pp. 10-11. Legally Free Computer Books, US (2003)
- [11] Sadaghiani, H. R., & Pollock, S. J.: Quantum Mechanics Concept Assessment: Development And Validation Study. Physical Review Special Topics-Physics Education Research. Vol. 11(1), 010110. (2015)
- [12] Sugiyono.: Metode Penelitian Kuantitatif dan R & D. Alfabeta, Bandung (2011)
- [13] Muslim, Suhandi, A., & Nugraha, G.: Development of Reasoning Test Instruments Based on TIMSS Framework for Measuring Reasoning Ability of Senior High School Student on the Physics Concept. In Journal of Physics: Conference Series. Vol. 812, No. 1, p. 012108. IOP Publishing (2017)
- [14] Yasuda, J. I., & Taniguchi, M. A.: Validating Two Questions In The Force Concept Inventory With Subquestions. Physical Review Special Topics-Physics Education Research. Vol. 9(1), 010113. (2013)
- [15] Yunita, W.: Pengembangan Tes Fisika SMA Kelas X Semester Ganjil. Skripsi: Fakultas Keguruan dan Ilmu Pendidikan, UNS. (2012)