Development of Electrochemistry Testlet for Measuring Higher Order Thinking Abilities for Student in Vocational High School

Indah Damayanti[1,*], Dr. Muhammad Masykuri[2], Dr. Sri Yamtinah[3]

[1]SMK SMTI Yogyakarta, Jl. Kusumanegara No 3 Yogyakarta, *email: indie.enkui@gmail.com
[2,3]Science Education Magister Department, Teacher Training and Education Faculty, Sebelas Maret University Surakarta

Abstract

This study aims to develop assessment instrument testlet that has validity, reliability, level of difficulty, and distractor that has good criteria, and to develop testlet instrument in electrochemistry subject matter that can be measure student's Higher Order Thinking Abilities. Development of testlet assessment instrument for measuring student's Higher Order Thinking Skills procedure has 10 steps, those are: 1) research and information collecting, 2) planning, 3) developing preliminary form of product, 4) testing preliminary field, 5) revising main product, 6) testing main field, 7) revising operational product, 8) testing operational field, 9) revising final product, 10) dissemination. The subjects of this research are 352 students of SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK SMAK Padang, and SMK SMAK Makassar. Validity of this research uses Aiken formula with 7 experts. Assessment method of the testlet instrument uses Graded Response Model (GRM).

The result of testlet assessment instrument is fulfill as a good instrument which has validity more than 0.76 (valid), precentage of reliability in main field test is 0.889 (high) and precentage of reliability in operational field test is 0.802 (high), discriminator power percentage is 50% enough and 50% good. Level difficulty percentage is 8.33% difficult, 63.89% enough, 11.11% easy, and 16.67% very easy. Testlet assessment instrument could measured student's Higher Order Thinking Abilities. The student's higher order thinking abilities percentage is analysis ability 70.05%; evaluating ability 54.23%; creating ability 42.66%; critical thinking ability 54.59%; and logical reasoning ability 36.41% at testing operational field.

Keywords: Research and Development, Testlet, Assessment Instrument, Higher Order Thinking Abilities, Electrochemistry

Introduction

The national education system is a whole components education the related in an integrated way to achieve national education targets. National education targets become the main parameters to formulate national education standards. National education standards are the basis in planning, implementation, and supervision education in the context of creating national education quality. National education standards consists of eight standard, one of which is assessment standards aims to ensure:

1. Planning assessing in accordance with competence to be accomplished and based on the principle of judgment;
2. The implementation of the assessment professionally , open , educational , effective , efficient , and as appropriate to the context social and cultural; and
3. Reporting the assessment results of the objectively, accountable, and informative

To do an assessment, needed an instrument in accordance with competence to be measured, effective, efficient, educational, objective and accountable. For vocational students, apply category can controlled, because students have took some getting used to apply or using a procedure.But, for the ability higher order thinking skills, the students must trained and took some getting used. The students training to think a high degree, one of which may be through assessment activities study results students.With an instrument assessment that can measure the capacity of higher order thinking skills, students will be called to analyze, evaluate, create, think critically, and logical reasoning.

According to King, Goodson, and Rohani (2004 : 1) higher order thinking skills is the ability where students able to think critically , logical , reflective , metacognitive and creative. Good & Brophy (Gani, Auliah, & Faika, 2011: 4) thinking is an exercise and apply the cognitive ability, as posing, focus to answering the question, searching memory, processing information and evaluation potensial solution to problems.

The results of research conducted by Gani, Auliah, and Faika (2011: 8), shows that higher order thinking skills is the key in learning success. The research it's known that the students higher order thinking skills still low especially in a complex problem. The results of research by Fischer, Bol, dan Pribesh (2011: 5) show that higher order thinking skills rarely to apply in small calss, as a class to prepare the students will go to college.

Based on the result of teacher questionnare and discuss with four Physical Chemistry teachers from SMK SMTI Yogyakarta, SMK SMAK Bogor, SMK

SMAK Padang, and SMK SMAK Makassar, it's known that the vocational students higher order thinking skills no maximal exercising, so that necessary instrument that can exercise higher order thinking skills of students.

Based on the result of student questionnare is known 83.33% from 120 students said need to be developed an instrument to measure higher order thinking skills, 77.5% of students said the existed instrument is essay, 63.33% of students said guess the answer when not capable answering the difficult question at multiple choice questions. Thereby, it's known that the students will apply their thinking skills if find the essay question.

According to Susongko (2010: 270), essay questions more informative than multiple choice questions concerned the capability of students. Essay question can shows the students logical ability. The students which aren't competence about the topics will not guessing the answer of examination. Essay question had a lot of benefit, but it's scorring need to much of time, and appearance of the teacher subjectivity eventhough using scorring criteria and using more than one *rater*.

The alternative of assessment instrument is *testlet*. Testlet is an question group that related to a given topic that developed into one of unity and contains a number of steps that has been specified beforehand and that can be followed by participants. According to Wainer, Sireci, and Thissen (1991: 187) *testlet* including into test type which will produce more than of one response, further testlet it has a response that hierarchical in relation to knowledge (construct) to be measured.

Testlet having many advantage, scorring system is politomus (not score dichotomy like the multi choice question), scorring not take too long of time, and scorring system is hierarchy, where is the first question the basic of second question and so the second question is the basic of third question. So the students' higher order thinking skills are evident in how far students were able to solve every stage about, from the first to third on about.

According to Yamtinah , Haryono, and Martini (2014: 4) an item of question on testlet consisting of several question about a supporter who is dependent. When

the first question is answered wrong by students , they will not be able to answer right about at the next stage (question) to a whole. This will affect the score obtained students.

The hierarchy questions will affect the score, so scorring use *Graded Respon Model* (GRM). De Ayala, Dodd, and Koch (Wahyuni, Yamtinah, & Utami, 2015: 228) GRM is one model developed by Samejima to scorring politomus item .With GRM, teachers can look at the thought processes students like on essay question, but scorring in the form of testlet more efficient because is objective and politomus .The process of scorring using GRM is as follows:

Table 1. Scorring criteria with *Graded Respon Model* (4 category)

Number	Aspects Assessment	Score
1	Students cannot finish step on the first stage properly	0
2	Students completed the first stage correctly , but cannot finish second stage	1
3	Students completed the first and second stage correctly, but cannot finish second stage	2
4	Students can solve the whole of step correctly	3

One example of an instrument for the testlet is as follows:

Stem:

Known the oxidation reduction reaction:

$Cl^- + MnO_4^- \longrightarrow Cl_2 + MnO_2$

Questions:

1. Oxidation state Mn change from ....
   a. +7 to +4
   b. +4 to +7
   c. 0 to -1
   d. -1 to 0
   e. +7 to +2
2. On it's reaction, each mole of MnO_4^- accept the electron ....
   a. 1 mole
   b. 2 mole
   c. 3 mole
   d. 5 mole
   e. 7 mole

3. Reaction equation from it's equation is ....
   a. $6Cl^- + 2MnO_4^- + 4H_2O \longrightarrow Cl_2 + 2\ MnO_2 + 8\ OH^-$
   b. $6Cl^- + 2MnO_4^- + 4H_2O \longrightarrow Cl_2 + 2\ MnO_2 + 8\ OH^-$
   c. $3Cl^- + MnO_4^- + 2H_2O \longrightarrow Cl_2 + MnO_2 + 4\ OH^-$
   d. $3Cl^- + 2MnO_4^- + 2H_2O \longrightarrow 3Cl_2 + 2\ MnO_2 + 4\ OH^-$
   e. $3Cl^- + 3MnO_4^- + 2H_2O \longrightarrow Cl_2 + MnO_2 + 4\ OH^-$

An instrument assessment the form of testlet will facilitate teachers to know students ability in resolving every stage about. Steps about will show how the students higher order thinking skills, the students ability to use skills of think analysis, evaluative, creative, critical, and logical reasoning.

Based on the result of student questionnare is known 66.67% from 120 students said electrochemistry need higher order thinking skills, because this topics need to analysis the electrolysis reaction in catoda and anoda. Another fourty students, have different answer; 16.66% of students said buffer solution; 15% of students said hydrolysis of salt; and 1.67% of students said stoichiometry need higher order thinking skills.

Based on the analysis result of information collecting the teachers and students, so the aims of developing assessment instrument testlet to knows:

1. The quality of testlet product, who in terms of: 1): the validity, language, and construct; 2) the quality of items of question which includes; reliability, discriminator power, level difficulty, and distractor.
2. The instrument testlet assessment in assessing or measure the higher order thinking skills of electrochemistry in vocational schools.

Material and Method

Model of development used in the development of an instrument assessment testlet for measuring higher order thinking skills this is research and development. Steps on the research is step research and development developed by Borg and Gall in 2007. The tenth the effort is: 1) research and information collecting, 2) planning, 3) develop preliminary from product, 4) preliminary field testing, 5)

main product revision, 6) main field testing, 7) operating product revision, 8) operational field testing, 9) final product revision, 10) dissemination and implementation.

Samples to this research is student will with fields of expertise chemical analysis derived from SMK SMTI Yogyakarta, SMK SMAK Bogor , SMK SMAK Padang, and SMK SMAK Makassar. Fourth this school shaded by the ministry of industry of the Republic of Indonesia .The schools was chosen to the distribution of data. Data processing in this report is written with using analysis quantitative. The quality of testlet product, who in terms of: the validity, language, and construct; and the quality of items of question which includes; reliability, discriminator power, level difficulty, and distractor.

At the development the preliminary draft assessment testlet an instrument for measuring higher order thinking skilla revised based on advice of the experts. Before tried out an instrument assessment testlet developed validated by the matter, the evaluation, the learning, and education experts by using Aiken formula. The criteria used is if $V \geq 0.76$ so the item is valid and to continue the analysis.

Result and Discussion

Data the assessment results of the validator to feasibility an instrument assessment testlet can be concluded that all aspects of a testlet has been declared valid or has been being used to the revision. The purpose of this discussion is to agree on indicators of higher order thinking skills as used in the development of a testlet and as an early stage of validation an instrument developed. This stage is stage develop of the preliminary form of the product.

Testing the quality of items about in terms of content was carried out by the 7 validators .This discussion produce input and suggestions on improvement about which has developed. Total of 36 items of questions about tested using Aiken formula show results more than 0.76; that is all about developed can be said in the valid.

The quality of an instrument developed can determined by means of testing empirical (empirical analysis). Testing empirical on the intrumen the assessment

is done with analysis grains about, consisting of testing reliability an instrument, discriminator power, difficulty level and the quality of distractor on the answers. Testing products an instrument testlet done as much as three times with different respondents every testing, the first phase the preliminary field and product revision, this stage is carried out by providing products an early assessment testlet on 16 students that is 4 students at random on each school .

This stage also intended to test an instrument, so it is expected that when the next stage the instrument can be read and be used clearly. The questionner at the trial this explains that the instrument a start given there are still mistake in number 10. This outcome be the foundation for revise an instrument early. A respond and advice students in general is about number 10, not complete so confuse to do, but generally an instrument given can be used as an instrument for measuring the higher order thinking skills.

Revise number 10 is;

Stem: Known the data of reduction potential from a volta cell:

Pb^{2+} (aq) | Pb (s) E° = -0,13 volt

Ag^{+} (aq) | Ag (s) E° = + 0,80 volt

10. The equation reaction of reduction from the data of reduction potential above is ....
    a. $Ag \longrightarrow Ag^{+} + e^{-}$
    b. $Ag^{+} + e^{-} \longrightarrow Ag$
    c. $Pb^{2+} + 2e^{-} \longrightarrow Pb$
    d. $Pb \longrightarrow Pb^{2+} + 2e^{-}$
    e. $Ag + Pb^{2+} \longrightarrow Ag^{+} + Pb$

become,

10. The equation reaction of reduction from an volta cell can be occured from the data of reduction potential above is ....
    a. $Ag \longrightarrow Ag^{+} + e^{-}$
    b. $Ag^{+} + e^{-} \longrightarrow Ag$
    c. $Pb^{2+} + 2e^{-} \longrightarrow Pb$
    d. $Pb \longrightarrow Pb^{2+} + 2e^{-}$

e. $Ag + Pb^{2+} \longrightarrow Ag^{+} + Pb$

The next testing is main field test and product revision then operational field test and the final product revision. At the main field test, an instrument that has undergone a revision of the previous given at a class (taken 20 students) from their respective school by the total number of 80 students while in the field operational, an instrument assessment testlet given to 2 classes in each school, and the number of students 256 students. The answer from students after do testlet electrochemical, analyzed in order to know the quality of grains or items of question. Analysis was about reliability, dicriminator power, the difficulty level and distractor. In addition, at this stage also analyzed the higher order thinking skills owned by students. The result of this analysis used as a basis to revise products. Analysis to be done manually, using software microsoft excel. Reliability at main field testing is 0,889 and 0,802 at operational field test.

The discriminator power is how good an item can to discriminate participants from construct measured. On the main field test, obtained discriminator power 30.56 % of items included in an enough category, 38.88 % of items in the good category, and 30.56 % of items in the very good category while in the operational field test 50 % of items included in an enough category and 50 % of items in the good category. A grains or items unable to discriminate students' ability with good if the distractor not functioning properly or about too easy.

Difficulty level items said the proportion of number of participants test the one who answers by the right items to all participants test. A number that can be given difficulty level items test called difficulty index denoted with letters P which means proportion.The higher value P, the more easily a items and the lower value P, more difficult. On main field test, difficult criteria as 13.88 %, medium as 63.89 %, easy as 16.67 % and very easy as 5.56 % while in operational field test difficult criteria 8.33 %, medium as 63.89 %, easy as 11.11 %, and very easy increased 16.67 %. The more items in very easy level, so discriminator power will go down, due either students who have high skill and low, will answer correctly. It caused between students who have the ability of high and low difficult to

discriminate. The comparative difficulty level on the main and operational field can be seen from Fig. 1.

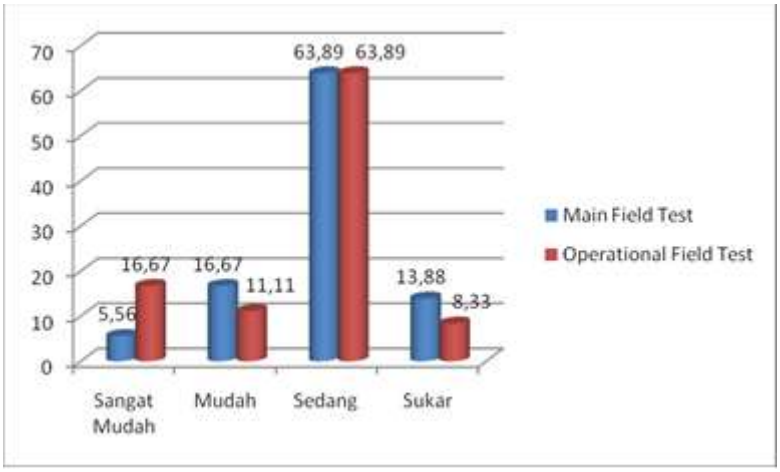

Fig. 1. The comparative of difficulty level

In Fig 1 the percentage the difficult items and easy items is down on the operational field test, but the percentage the very easy items surged to a significant. The more items in the category of very easy, then discriminator power item will go down, due either students who have high skill and low, will answer correctly.

The next analysis of item is a distractor. Distractor can be said good if the distractor has been having the power of attractiveness in such a way that participants tests (especially the ability of low) felt torn and hesitation so that in the end they choose the distractor as an answer right. The distractor said to have been functioning properly if chosen more than 5 % the total number of participants a test.

The main field test is carried out by providing an instrument assessment testlet to 80 students of 4 schools representative (each 20 students, from one a class). Data from this test will be analysis of item. The revision at this stage is based on two this. Data of analysis items that the less good of results will revised .Table 2 shows some option for which having index value of distractor 0.00. This indicates that option the answer not chosen by all participants test and shows that the distractor did not function properly. To testing next stage the distractor having the value 0.00 must be revised.

Table 2. Data of the poor quality of items

No item	Discriminator Power	Difficulty Level	Distractor (%)				
			A	B	C	D	E
1	0.230	0.925	92.50	5.00	2.50	**0.00**	**0.00**
2	0.282	0.913	3.75	91.25	3.75	1.25	**0.00**
4	0.403	0.850	85.00	8.75	2.50	**0.00**	3.75
6	0.468	0.825	2.50	82.50	10.00	5.00	**0.00**
7	0.247	0.888	88.75	8.75	1.25	1.25	**0.00**
8	0.413	0.313	65.00	31.25	**0.00**	3.75	**0.00**
9	0.307	0.838	83.75	2.50	8.75	**0.00**	5.00
14	0.629	0.300	30.00	8.75	1.25	60.00	**0.00**
15	0.222	0.450	30.00	45.00	6.25	18.75	**0.00**
18	0.645	0.400	**0.00**	40.00	20.00	28.75	11.25
32	0.574	0.413	41.25	33.75	**0.00**	17.50	7.50
34	0.450	0.338	32.50	**0.00**	1.25	33.75	32.50

Table 2 shows that there are six numbers, the number 1, 2, 4, 6, 7, and 9 distractor D or E no voters at all. In some items which is very easy and having the answer A or B, possible students have struggled to find another option, so the distractor overlooked or not read again by students. Alternative solutions to phenomena like this is the answer (when no number) could be put in option last or option E.

In general, an assessment instrument only has a function to measures the cognitive test, but the development of testlet based on an indicator cognitive adapted to higher order thinking skills indicators. An instrument testlet developed consist of 12 stem, every stem has three supporter questions who is hierarchical, it means students not be able to answer number 3 true without through number 2, and students not be able to said about number 2 with true without through number 1. The hierarchical questions as it allows for the implementation of the GRM model. The application of GRM model in this testlet is teacher or users can understand the extent of the process knowledge possessed students, who in this research is the students higher order thinking skills. Every about made tiered, first question of any stem is about the fundamental concept of who should be easy and understood students

Each item made a series of steps, first question of any stem is about the fundamental concept of who should be easy and understood students. This is in accordance with the opinions Susongko (2010: 287); Yamtinah, Haryono, and

Martini (2014: 9) found that the participants score is a summary a score of every step. Distribution number based on an higher order thinking skills indicator is presented in table 3. Results of the analysis of the higher order thinking skills of students at the main field test is presented in table 4.

Table 3. Distribution number based on higher order thinking skills indicator

Indicator	Number of item
Analyse	1, 4, 5, 6, 7, 8, 9, 15, 22, 23, 24, 34
Evaluate	2, 3, 19, 20, 31, 35
Create	18, 25, 26, 27, 30
Critical Thinking	10, 11, 12, 13, 14, 16, 17, 28
Logical Reasoning	21, 29, 32, 33, 36

Table 4. Completeness of higher order thinking skills indicator at main field test

Indicator	Percentage of completeness (%)				Total
	SMK SMTI Yogyakarta	SMK SMAK Bogor	SMK SMAK Padang	SMK SMAK Makassar	
Analyse	49.17	75.83	42.08	77.50	61.15
Evaluate	48.33	65.00	30.83	65.83	52.50
Create	21.00	57.00	0.00	58.00	34.25
Critical Thinking	23.13	77.50	12.50	56.88	42.50
Logical Reasoning	50.00	53.00	0.00	46.00	37.25

Table 4 shows the percentage of analyse indicator is the highest percentage in 3 three schools, exceptually SMK SMTI Yogyakarta. It's show that students ability to analyse a case and explain reasons used to connects the part with each other has been good enough. The lowest percentage of higher order thinking skills indicators is logical reasoning indicator at three schools are SMAK Bogor, SMAK Padang, and SMAK Makassar. But for SMK SMTI Yogyakarta's students, percentage of logical reasoning indicator is the highest percentage. Logical reasoning ability is the ability to explain a fact, the truth in accordance with reasoning.

Table 5. Completeness of higher order thinking skills indicator at operational field test

Indicator	Percentage of completeness (%)				Total
	SMK SMTI Yogyakarta	SMK SMAK Bogor	SMK SMAK Padang	SMK SMAK Makassar	
Analyse	62.37	81.38	53.13	83.33	70.05
Evaluate	48.96	61.46	44.27	62.24	54.23
Create	23.75	60.63	20.63	65.63	42.66
Critical Thinking	31.25	79.88	41.60	65.63	54.59
Logical Reasoning	43.43	38.44	12.19	51.56	36.41

Table 5 shows the result is no different far by testing to the earlier stages. Indicators of higher order thinking skilla the most controlled the students were indicators analyze, while an indicator that the lowest percentage is logical reasoning. An other indicator, in each school give comparison that is almost the same of testing before.This may be due to the quality of items that the inclined to each.The difficulty level also not changed. Analyze indicator more likely to done than indicators the higher oreder thinking as other. Analyze indicator there in first question in each stem, so this is more easy than 2nd and 3rd question.

The result of testing can be used to categorize the students based on the higher order thinking skills of owned. In this study, the students score categorize into the 3 level; high, medium and low to the higher order thinking skills of owned. The students said high when >24, medium when 13-24, and low when < 13. The comparison of students score at main field test is presented in Fig 2 and the comparison of students score at operational field test is presented in Fig 3.

Fig. 3 shows, the students of SMK SMTI Yogyakarta dominant in a low score, while for students in SMK SMAK Bogor dominant in a medium score. Students SMK SMAK Padang dominant in a low score and for students SMK SMAK Makassar have the same percentage of high score and medium score.

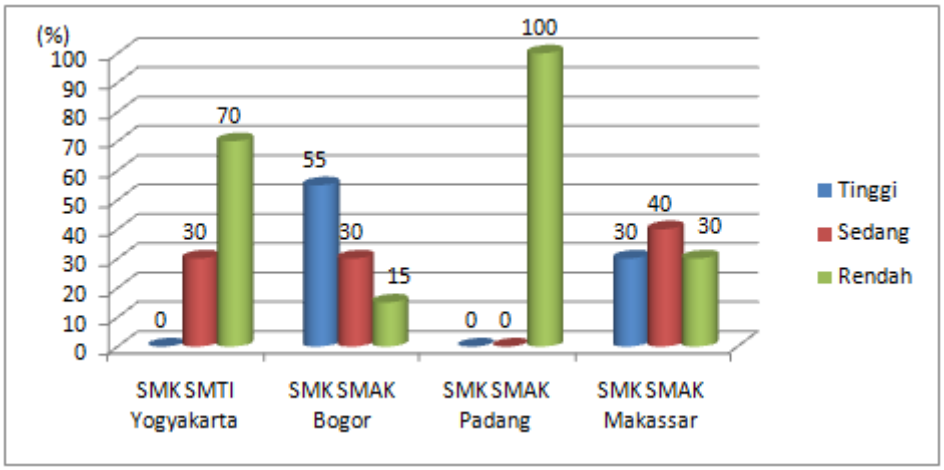

Fig 2. The comparative of students score at main field test

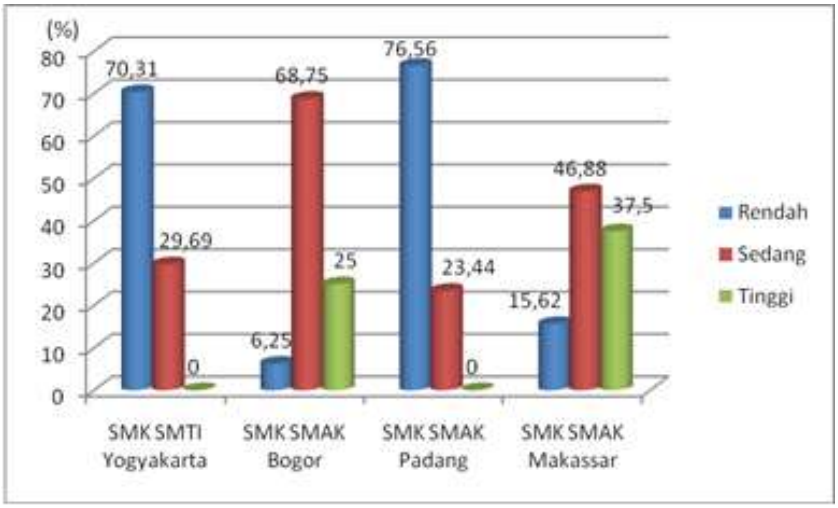

Fig 3. The comparative of students score at operational field test

In this study, the ability an instrument assessment testlet in measuring the higher order thinking skills of matter electrochemical, analyzed based on the percentage a score these students on each indicators the higher order thinking skills of students. In table 5, show the percentage of students completeness on each indicators the higher order thinking skills, sample 4 students that each a student from each school.

Table 6. Completeness of higher order thinking skills indicator (*sample*) at operational field test

Indicator	Completeness (%)			
	Student A (SMK SMTI Yogyakarta)	**Student B (SMK SMAK Bogor)**	**Student C (SMK SMAK Padang)**	**Student D (SMK SMAK Makassar)**
Analyse	75.00	100.00	58.33	91.67
Evaluate	83.33	66.67	33.33	100.00
Create	80.00	100.00	0.00	80.00
Critical Thinking	12.50	87.50	12.50	75.00
Logical Reasoning	100.00	80.00	0.00	60.00

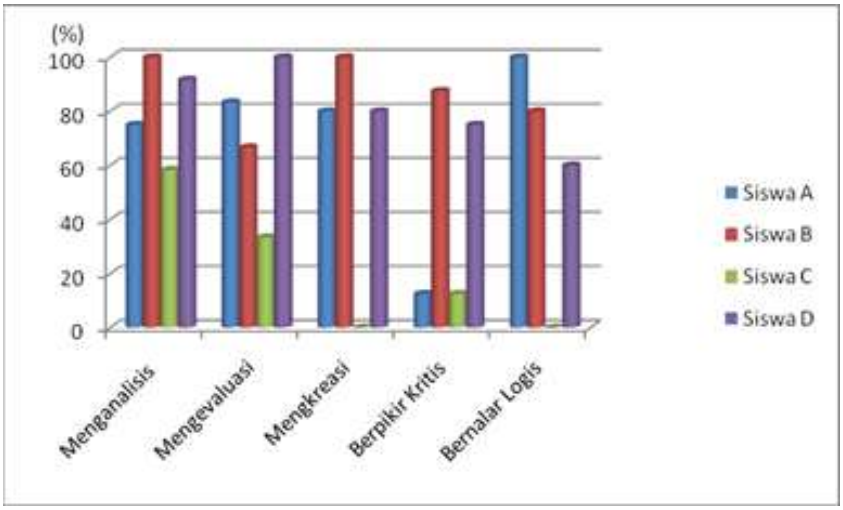

Fig 4. The comparative of students score at main field test

In table 5 and figure 4 it can be seen that students A need to increase in critical thinking, because the percentage most small compared with the ability analyze, evaluate, create, and logical reasoning, while students B need to increase on the ability of evaluate. Students C need to increase in five indicators the higher order thinking skills, namely the ability analyze, evaluate, create, critical thinking, and logical reasoning. Students D, need to increase on the ability of logical reasoning. Students A prominent on the ability of logical reasoning, while students B prominent on the ability of to analyse and create, and students logical reasoning prominent on the ability of evaluate.

Conclusion

Based on the research result and discussion, can be concluded that:

1. Instruments testlet developed announced eligible and have the criteria as a good items of question with the results of contents validity using a formula Aiken, obtained the material developed has the validity more than 0,78 which is said to be valid, reliability on the main field test is 0,889 and 0,802 on operational field test who could be classified as very high, having the discriminator power with the percentage of 50 % enough and 50% in good the operational field test, as well as having a difficulty level 8.33 % very difficult, 63.89 % medium , 11.11 % easy, and 16.67 % very easy on operational field test.
2. Testlet instrument for measuring the higher order thinking skills of matter electrochemistry developed capable of measures the higher order thinking skills of students with the percentage of students ability to analyse of 70.05 %, evaluate 54.23 %; create 42.66 %, critical thinking 54.59 %, and logical reasoning 36.41 % by the operational field test.

References

Aiken, L. (1985). Three Coefficient for Analyzing The Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45, 131-142.

Borg, W.R., & Gall, M.D. (2007). *Educational research*. New York: Longman, Inc

Fischer C., Bol L., & Pribesh S. (2011). An Investigation og Higher Order Thinking Skills in Smaller Learning Community Social Studies Classrooms. *American Secondary Education Spring Journals*. 39:2.

King F.J., Goodson L., & Rohani F. (2004). Higher Order Thinking Skills, Definition, Teaching Strategies, Assessment.

Susongko, P. (2010). Testlet Sebagai Alternatif Pengembangan Bentuk Tes Prestasi Belajar pada Pembelajaran Kimia di Sekolah. *Jurnal Penelitian dan Evaluasi Pendidikan*, 14(2) 269-288.

Gani T., Auliah A., Faika S., (2011). Penguasaan Pengetahuan Deklaratif dan Kemampuan Berpikir Tingkat Tinggi Mahasiswa Prodi. Pendidikan Kimia. *J. Chemica.* 12:1-9.

Shidiq, Ari Syahidul. (2016). Pengembangan Instrumen Penilaian Testlet untuk Mengukur Keterampilan Proses Sains pada Materi Hidrolisis Garam untuk Siswa Kelas XI SMA/MA. *Tesis S2 Fakultas Keguruan dan Ilmu Pendidikan Universitas Sebelas Maret*. Surakarta. (Unpublished).

Wahyuni, I.T., Yamtinah S., & Utami B. (2015). Pengembangan Instrumen Pendeteksi Kesulitan Belajar Kimia Kelas X Menggunakan Model Testlet. *Jurnal Pendidikan Kimia*, 4(4) 222-231.

Wainer, H, Sireci, S. G & Thissen, D. (1991). Differential Testlets Functioning: Definition and Detecting. *Journal of Education Measurement*, 24, 185-201.

Yamtinah S., Haryono, & Martini, K.S. (2014). Profil Individu Peserta Didik Pelengkap Tes Jenis Testlet Sebagai Alternatif Pendeteksi Kesulitan Belajar Kimia. *Jurnal Profesi Pendidik*, 1(1) 1-10.