



Research Article

# The Effect of Multiple Choice Scoring Methods and Risk Taking Attitude toward Chemistry Learning Outcomes

Awaluddin Tjalla and Sari Fitriani

Jakarta State University, Indonesia

Correspondence should be addressed to: Awaluddin Tjalla; awaluddin.tjalla@yahoo.com

Received date: 1 March 2017; Accepted date: 22 May 2017; Published date: 8 September 2017

Copyright © 2017. Awaluddin Tjalla and Sari Fitriani. Distributed under Creative Commons CC-BY 4.0

## Abstract

The objective of this study is to investigate the effect of multiple choice scoring methods Number Right Elimination Testing (NRET) and Formula Scoring (FS) toward chemistry learning outcomes by considering students' risk taking attitude. The study was conducted in SMAN 13 Kota Bekasi. Samples of this study were 40 students of XI grade in Science Program Class during the second semester of the academic year 2015/ 2016. The samples were selected by simple random sampling technique. The method of this study used an experimental method. While the design of the study used treatment by level 2x2. Data collection was conducted by measuring students' risk taking attitude and students' chemistry learning outcomes. The hypotheses of the study were tested using two way analysis of variance (ANOVA). The result showed that: (1) There were differences in chemistry learning outcomes between students who were tested with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, (2) There was an interaction effect between scoring methods on multiple choice tests and risk taking attitude toward chemistry learning outcomes, (3) There were differences in chemistry learning outcomes between students who are categorized as risk seeking and tested with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, and (4) There were differences in chemistry learning outcomes between students who are categorized as risk averse and tested with multiple choice tests using Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method.

**Keywords:** Multiple Choice Scoring Method; NRET; FS; Risk Taking Attitude

## Introduction

Chemistry is the foundation of the natural sciences such as, medical science,

pharmacy, geology, engineering, and others. Therefore, chemistry is an important part of science that has a major role to play in the development of a nation.

But until recently, studying chemistry in high school was considered difficult by students, it is so because of the characteristics of chemistry itself, those are: chemistry teaching materials that are abstract; chemistry is a simplification of the actual situation; chemistry teaching materials are sequential and evolving at a fast rate; the objective of learning chemistry is not only to solve problems but also to learn the descriptions of the facts of chemistry, the rules of chemistry, the terms of chemistry; and materials studied in chemistry are many (Middlecamp & Kean, 1985: 9). The difficulty of studying chemistry impacts on students' learning

outcomes in senior high school, which is still relatively low.

Based on observations that were conducted at SMAN 13 Kota Bekasi, It was obvious that the results of studying chemistry in senior high school are still low in terms of achieving minimum criteria score (KKM) and the acquisition of the national examination score. Average acquisition of the national examination score of chemistry for the last three years at SMAN 13 Kota Bekasi is still relatively low, which falls in the classification C and D, where classification C stands for moderate level of achievement, and classification D stands for low level of achievement.

**Table 1: Average of the National Examination Score of Chemistry at SMAN 13 Kota Bekasi for the Last Three Years**

Academic Year	2012/ 2013	2013/ 2014	2014/ 2015
Classification	C	D	C
Average	6,45	5,25	6,77
Lowest	4,00	3,25	4,00
Highest	9,25	8,50	9,00
Standard Deviation	1,26	1,24	1,15

(Source: List Collective of National Exam Results (DKHUN) SMAN 13 Kota Bekasi)

Learning outcomes are a reflection of the success of learning programs in schools, where learning is itself a series of activities that constitute the process of the interaction of learners with educators and learning resources in a learning environment (Suryabrata, 2012: 233). Learning is a deliberate effort by educators to convey science, organize and create a system environment with a variety of methods, so that learners can perform learning activities effectively and efficiently to achieve optimal results. The effectiveness of the learning process can be seen from the achievement of learning objectives that are reflected in students' learning outcomes obtained after following a series of specific learning processes. The achievement can only be known if an educator conducts an evaluation of learning outcomes of students. Evaluation is then used as a benchmark to determine

how far the level of competence the learners achieved of the material already learned.

The main role of the evaluation is to determine the level of achievement of the education objectives and learning objectives that have been determined by looking at the results that have been able to be achieved by learners. Evaluation is a continuous process. Evaluation can be done before or after the learning process, even the evaluation should also be carried out during the learning process.

During the learning process, the evaluation can be done through formative evaluation procedure. Where the formative evaluation is an evaluation carried out in the middle or at the time of the learning process, which is carried out on each unit of learning programs can be completed by learners, in order to determine the extent

to which learners "have been established", in accordance with the learning objectives that have been determined (Sudijono, 2007: 23). Formative evaluation is a series of evaluations based on classroom assessment, which is an integral part of the learning process, is carried out as the process of data collection, utilization of thorough information about the learning results obtained by the students to determine the level of achievement and mastery of competences which have been determined by the curriculum and as corrective feedback for learning process (Sanjaya, 2008: 350).

Less satisfying results of studying chemistry in senior high school can be caused by several factors. Factors that influence learning outcomes can be either internal or external (Suryabrata, 2013: 233). One factor that is believed to be able to influence the results of studying chemistry is the selection of the test item format in evaluation program that is applied by teachers in the classroom, especially in the formative evaluation. Formative evaluation carried out, not merely provides score as a result of the measurement process, but it gives the meaning of the score achieved by students, and at the stage of reflection, teachers can motivate students for further improvement of the learning process so that the achievement of competencies as defined in learning objectives can be seen in the results obtained by students.

Multiple choice test item format is the most common used test format in various educational evaluation programs, it is because multiple choice can be scored easily and quickly, objectively, analyzed more easily, may include extensive material on a test, can measure the ability of a variety from the simplest to the most complex. In the formative evaluation program, a multiple choice test is generally scored with conventional scoring methods called number right scoring method (NR). In conventional scoring method (NR scoring method), each test score is the sum of the item scores for a given examinee, and the examinee is awarded one point for the correct item response and zero for any

other response (Crocker & Algina, 1986: 399). The main concern in this scoring method is that students can answer correctly by guessing (Choppin, 1988: 384-386). The number of correct answers on a conventional scoring method can consist of two things; the number of questions where the students actually knew the answer, and the number of questions to which the students guess the answer correctly. Sometimes, students only know the part of the answer or answered uncertainly, it is known as partial knowledge (Coombs, Milholland, Womer, 1956: 36). Therefore, when using this conventional scoring method, the teacher cannot distinguish whether the student answers correctly based on the knowledge and ability or based on guesses. Guesses that benefit will boost students' scores, thus, causing an overestimate of the ability of the student.

Using NR scoring method in multiple choice test, not only partial knowledge of students is not credited, but teachers also cannot diagnose students' misunderstanding and lack of understanding in order to provide informative feedback to facilitate students' continuous learning (Lau, et al, 2011: 99). Additionally, Kulhavey, et al., said that students spend less time learning for multiple choice tests relative to an essay tests (Roediger & Marsh, 2005: 1155). Sudjana (2009: 43) argues objective test has been considered as an easy test to take and prepare for the exam at school (43). It shows that the multiple choice test with conventional scoring methods do less stimulate students' motivation, hence influence the learning outcomes of students.

The weaknesses of multiple choice tests mentioned make chemistry teachers in senior high school tend to only use essay test format than multiple choice test format in formative evaluation. But then another problem arises as the Indonesian Government Regulation number 74 of 2008 concerning teacher explained that the workload of teachers must be at least 24-hour face to face and at most 40 hours of face to face in one week at one or more units of educational institutions that have a

license from the government or local government. It means chemistry teachers must teach about 240 to 400 students every week (public schools have 40 students per class). Based on that regulation, when teachers always implement essay test format in formative evaluation, then teachers will be very busy correcting test answer sheets in every single office day, considering essay test takes longer and is difficult for correcting than the multiple choice test. Furthermore, if the tests' correction takes so long, then the feedback expected from the formative evaluation could become irrelevant for the students, because of the large time gap between the tests and feedback.

Occurrence of problems in the evaluation learning program, as described earlier, ultimately drives the born of alternative scoring methods of multiple choice test that can manage an answer (response) from test participants with more complex way by predicting guessing tendencies and detecting partial knowledge. Over time, and with the proper training in the use of the alternative scoring methods in multiple choice tests, the alternative scoring methods can improve students' motivation and academic grades, because after all, the ultimate goal of the test is to provide adequate information for teachers and students that can be used in efforts to enhance the process of teaching and learning.

There are so many alternative scoring methods of multiple choice test which have been introduced by experts, but this research will only focus on the study of Formula Scoring (FS) method and the Right Number Elimination Testing (NRET) scoring method, where both scoring methods are equally applying the penalty system on the wrong answer. Penalty system is the traditional approach used in an attempt to reduce guesses on the answers to multiple choice tests. The underlying logic of this approach is to prevent the test participants from earning points that should not be accepted (Crocker & Algina, 1986: 399-400). A penalty system in scoring multiple choice tests is expected to make students be more careful in taking the test and avoid guessing. Penalty here is a sentence reduction of the value of the score (Naga, 2013: 88). When students are already aware that guesses would risk reducing their exam scores, then students will prepare better for the test and modify their learning strategies so that they can answer the test correctly without having to guess. Different with FS method, NRET is also equipped with a credit/ score on the partial ability of students in addition to a penalty system, so that NRET is able to detect the partial knowledge/ abilities and students' misconceptions.

The mathematical formula for the FS method is as follows (Frary, 1988: 33).

$$FS = R - \frac{W}{(C - 1)}$$

In which

FS = "corrected" or formula score

R = number of items answered right

W = number of items answered wrong

C = number of choices per item (same for all items)

Table below contains the test instruction, scoring guides and scoring taxonomy for NRET. (Lau, et al, 2011: 100-106).

**Table 2: NRET Test Instruction, Scoring Guides and Scoring Taxonomy**

<b><u>NRET Test Instruction:</u></b>	
<p><b>Must choose ONE option as the ANSWER by using "✓ CORRECT."</b>  <b>ELIMINATE option(s) that you are SURE ARE NOT THE ANSWER by using "X WRONG."</b>  <b>USE "? NOT SURE" if you are NOT SURE of an option.</b>  <b>You have the flexibility to choose NONE (0), ONE (1), TWO (2), or THREE (3) "X WRONG," or "? NOT SURE."</b></p>	
<b><u>NRET Scoring Guides:</u></b>	
<p><b>ONE (1) point awarded if the option with "✓ CORRECT" is the correct answer.</b>  <b>ONE (1) point awarded for each option eliminated correctly with "X WRONG".</b>  <b>A PENALTY of 3 points deducted if the correct answer is eliminated with "X WRONG."</b>  <b>Your score will range from -3 to 4.</b></p>	
<b><u>NRET Scoring Taxonomy:</u></b>	
<b>Score</b>	<b>Category</b>
-3	<i>Full misconception (FM)</i>
-2	<i>Partial misconception (PM)</i>
-1	
0	<i>Absence of knowledge (AK)</i>
1	<i>Partial knowledge (PK)</i>
2	
3	
4	<i>full knowledge (FK)</i>

In most multiple-choice tests with the scoring method using penalty for wrong answer, the test taker must decide whether to answer a question (and risk losing points if the answer is wrong) or not answer the question. There, the test participants asked to make decisions under uncertainty. In situations such as tests, the rule in scoring the test can be considered as something risky. In the implementation of the multiple choice test, not only mastery of the subject matter (cognitive ability) is shown by students, but the tendency of the attitude of risk-taking, self-confidence in deciding which option they choose as the answer, as well as the strategies and attitudes of others can come into play as well when students face uncertainty, especially when students work in a limited time. In the process of learning that applies formative evaluation system using multiple choice tests, the psychological state, especially the attitude of risk-taking propensity (risk-taking

attitude) of the students, can affect students' overall learning strategy that ultimately affects the study results.

It is obvious that people differ in the way they resolve work-related or personal decisions that involve risk and uncertainty (Weber, Blais, Betz, 2002: 263). According to the tendencies to take a risk, people are categorized as risk seeking, risk neutral and risk averse (Weber, Blais, Betz, 2002: 263-290). The relationship between the scoring method of multiple choice test and risk taking attitude toward chemistry learning outcomes is important for further investigation.

### **Objectives and Significance of Study**

#### ***Objectives***

The general objective of this study was to determine: (1) the differences in chemistry learning outcomes among students who

were given a multiple-choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, (2) the interaction between multiple choice test scoring methods (Number Right Elimination testing and Formula Scoring) and students' risk taking attitude towards chemistry learning outcomes, (3) the differences in chemistry learning outcomes among students categorized as risk seeking given a multiple choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method, and (4) the differences in chemistry learning outcomes among students categorized as risk averse given a multiple choice test with Number Right Elimination Testing (NRET) scoring method and Formula Scoring (FS) method.

### **Significance of Study**

The results of this study are expected to be useful to: (1) provide insight to the teacher about the various scoring methods of multiple choice test that can be used in a formative evaluation in order to stimulate students' motivation to learn and improve student learning outcomes; (2) provide insight on risk taking attitude of students in relation to learning outcomes, especially in the use of various scoring method in multiple choice test; (3) serve as the basis for strategic decision-makers in education, especially the policies relating to the educational evaluation; (4) be used as empirical foundation for future researchers, both in studying the evaluation of education as well as the characteristics of individual students.

### **Method**

#### ***Research Design***

This research used experimental methods with design treatment by level (2x2). The treatment variable in this study is scoring methods on multiple choice tests, while attribute variable is a risk taking attitude that consists of two category, those are risk seeking and risk averse. The dependent variable in this study is chemistry learning outcomes.

Experiments carried out by treating such formative multiple choice test with NRET scoring method and FS scoring method are carried out systematically as many as four times. Wherein each formative test done every sub topic in the syllabus of learning chemistry completed taught by the teacher. After four formative tests are completed, the students are given an achievement test. To avoid bias in research, it is necessary to control the internal validity and external validity of the experiment.

### **Population and Sample**

The population in this study was 155 students of SMAN 13 Kota Bekasi in 11<sup>th</sup> (XI) grade (students are typically 16-17 years of age) of Natural Sciences (IPA) major. Research samples were taken by using simple random sampling technique. Successively conducted by: (1) choose two classes of the four classes XI IPA at SMAN 13 Kota Bekasi with the simple random sampling technique, using a lottery system (Gulo, 2005: 84); (2) The elected class was given treatment formative multiple choice test with NRET scoring method and one other class treated with formative multiple choice test with FS scoring method ; (3) Conducting risk taking attitude tests simultaneously to all students in the two classes XI IPA elected to obtain information risk taking attitude of students, using risk taking attitude instrument; (4) The risk taking attitude scores of the two classes are sorted from largest to smallest; (5) students who have an average score on a risk taking attitude instrument of greater than 0.5 standard deviations above the average are categorized as risk seeking, while students who have an average score on a risk taking attitude instrument of less than 0.5 standard deviations below the average are categorized as risk averse, the number of samples were selected in each cell in accordance with the study design is 10 students, so the total number of samples is 40 students.

### **Instruments and Data Collection Procedure**

In this research, there are two kinds of data collected through the research

instruments; those are the data of students' chemistry learning outcomes and data of students' risk taking attitude. Instrument of chemistry achievement test was used to measure student learning outcomes after getting treatment. Chemistry achievement test developed by researchers consists of 28 multiple choice items for chemistry topic of hydrolysis salt solution and buffer, internal consistency reliability of this instrument is 0.83. For the risk taking attitude instrument, researchers adapted from A Domain-specific Risk-attitude Scale (DOSPERT scale) developed by Weber, Blais and Betz (Weber, Blais, Betz, 2002: 263-290), based on the discussion with some experts in educational evaluation and educational psychology, the items in each domain were modified to measure students' risk taking attitude specifically on the situation of learning in schools especially learning chemistry, consists of 35 items with internal consistency reliability coefficient of 0.89.

### Data Analysis Procedures

Normality of the data was tested using the Lilliefors test, while the homogeneity of variance was tested using Fisher and Bartlett tests (Witte, 1985: 225-238). The hypothesis was tested using two ways analysis of variance (ANOVA) followed with simple effect test using Dunnett's test (Howell, 2007: 303 – 309).

### Statement of the Problem

To determine the effect of multiple choice scoring methods and risk taking attitude toward learning outcomes, In this study, the following questions were raised:

Is there a difference in chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method ( $A_1$ ) and FS scoring method ( $A_2$ ) ?  
Is there a significant interaction between multiple choice test scoring method (NRET and FS) and students' risk taking attitude towards chemistry learning outcomes?

Is there a difference in chemistry learning outcomes among students categorized as risk seeking given a multiple-choice test with NRET scoring method ( $A_1B_1$ ) and FS scoring method ( $A_2B_1$ )?

Is there a difference in chemistry learning outcomes among students categorized as risk averse given a multiple choice test with NRET scoring method ( $A_1B_2$ ) and FS scoring method ( $A_2B_2$ )?

### Results

Based on the statement of the problem and the theoretical framework, the research hypotheses were formulated:

Chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method ( $A_1$ ) are higher than chemistry learning outcomes among students who were given a multiple-choice test with FS scoring method ( $A_2$ ).

There is a significant interaction between multiple choice test scoring method (NRET and FS) and students' risk taking attitude towards chemistry learning outcomes.

Chemistry learning outcomes among students categorized as risk seeking given a multiple-choice test with NRET scoring method ( $A_1B_1$ ) are higher than chemistry learning outcomes among students categorized as risk seeking given a multiple-choice test with FS scoring method ( $A_2B_1$ ).

Chemistry learning outcomes among students categorized as risk averse given a multiple-choice test with FS scoring method ( $A_2B_2$ ) are higher than chemistry learning outcomes among students categorized as risk averse given a multiple-choice test with NRET scoring method ( $A_1B_2$ ).

The hypotheses in this research are tested using two ways analysis of variance (ANOVA), and the results can be seen in the following table:

**Table 3: Result of Two Way Anova for Hypothesis testing**

Source Variance	df	Sum of Squares	Mean Square	F	F <sub>crit</sub>	
					0,05	0,01
Scoring Methods (A)	1	319,225	319,225	8,497	4,11	7,39
Risk Taking Attitudes (B)	1	0,025	0,025	0,000665	4,11	7,39
Interaction (AxB)	1	1755,625	1755,625	46,730	4,11	7,39
Inter Group	3	2074,875	691,625	18,409	2,8	4,33
Within Group	36	1.352,500	37,56944	-	-	-
Reduction	39	3.427,375	-	-	-	-
Average	1	241025,625	-	-	-	-
Total	40	244453	-	-	-	-

As a consequence of the interaction, it is necessary to test the simple effect to examine the differences in chemistry learning outcomes among students categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method, and the

differences in chemistry learning outcomes among students categorized as risk averse given a multiple-choice test with NRET scoring method and FS scoring method. Dunnett's test results can be summarized in the following table:

**Table 4: Summary of Dunnett's Test**

Comparison Group	Dunnett's test		Conclusion
	t <sub>test</sub>	t <sub>crit (0,05)</sub>	
A <sub>1</sub> B <sub>1</sub> – A <sub>2</sub> B <sub>1</sub>	6,784	1,729	Reject H <sub>0</sub>
A <sub>1</sub> B <sub>2</sub> – A <sub>2</sub> B <sub>2</sub>	2,819	1,729	Reject H <sub>0</sub>

### Discussions

First, for the first research question, based on the results of two way ANOVA test, the value of  $F_{test} = 8.497$ , while the value of  $F_{crit}$  at significance  $\alpha = 0.05$  is 4.11. So  $F_{test} > F_{crit}$ , thus  $H_0$  is rejected. This means that there are significant differences in chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method and FS scoring method.

Based on the data collected, it is found that the average score of chemistry learning outcomes with students who were treated with multiple choice tests with NRET is higher than with students who were treated with multiple choice tests with FS. This finding is consistent with the different characteristics of the two scoring methods used in the study, even though both of

them use penalty as correction for guessing, but NRET is also able to provide opportunities and credit/score to the students' partial knowledge. Thus, the hypothesis put forward proved to be true, so the use of NRET scoring method in the multiple choice formative tests was found to be more efficient in estimating the ability of students and increasing the effectiveness of feedback in the learning process, which in turn positively affects student learning outcomes.

Second, for the second research question, obtained  $F_{test} = 46,73$ .  $F_{crit}$  at significance level  $\alpha = 0.05$  is 4,11. So  $F_{test} > F_{crit}$ , thus  $H_0$  is rejected. This means that there is a significant interaction between multiple choice test scoring method (NRET and FS) and students' risk taking attitude towards chemistry learning outcomes. This interaction can be interpreted as the fact



that the effect of multiple choice scoring method on formative evaluation depends on students' risk taking attitude.

Each student has a different risk taking attitude, there are some students who tend to avoid risk (risk averse) and some are more likely to seek risk (risk seeking). Students who have an attitude as risk averse always felt uncomfortable in such uncertain conditions in the test situation by using penalty system. While students who have the attitude as risk seeking are able to offset risks in a more relaxed way and are inclined to challenge the uncertain situation with the potential. Therefore, students with different risk taking attitude respond to multiple choice scoring methods in a different way as well, so the research data show there are different learning outcomes for each treatment given. This is confirmed by the results of two ways ANOVA test that showed a significant effect of interaction between multiple choice scoring methods and risk taking attitude towards chemistry learning outcomes.

Third, for the third research question, based on the result of simple effect analysis using Dunnett's test, for a comparison the average score of chemistry learning outcomes in students categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method, get the  $t_{test} = 6,789$  while  $t_{crit}$  value at significance level  $\alpha = 0,05$  is 1,729. So  $t_{test} > t_{crit}$ , thus  $H_0$  is rejected. This means that there are differences in chemistry learning outcomes among students categorized as risk seeking given a multiple-choice test with NRET scoring method and FS scoring method.

Students with a tendency to look for risk (risk seeking) remain a risk taker to guess the answers to multiple choice test although the test instructions have been said that the scoring rule implements a penalty for wrong answer, so guessing will not be tolerated. Students in this category looked at the FS scoring method as more neutral one than a threat. By contrast, when faced NRET scoring method, these students were facilitated and able to

mobilize all the potential for them to maximize the results of tests. This is supported by research data showing that the average score in chemistry learning outcomes of the risk seeking students who were treated with multiple choice tests with NRET was higher than with those who were treated with multiple choice tests with FS.

Fourth, based on the result of simple effect analysis using Dunnett's test, for a comparison the average score of chemistry learning outcomes in students who are categorized as risk averse given a multiple-choice test with NRET scoring method and FS scoring method, get the  $t_{test} = 2,819$  while  $t_{crit}$  at significance level  $\alpha = 0.05$  is 1.729. So  $t_{test} > t_{crit}$ , thus  $H_0$  is rejected. This means that there are differences in chemistry learning outcomes among students who are categorized as risk averse given a multiple choice test with NRET scoring method and FS scoring method.

Students with category risk averse respond with FS scoring method excessively, so these students trying to have maximal preparation before facing the test, because with the maximal preparation before tests made this group feel safe. Meanwhile, a system of penalties in NRET scoring method makes students with this category still feel threatened, so these students are less able to take advantage of the opportunity to maximize their score on NRET scoring method. On the other hand, the provision of credit/ score on the students' partial knowledge makes students with this category look at the test situation using NRET safer than the test situation using FS, this view makes students prepare less for the exam. This is supported by research data showing that the average score in chemistry learning outcomes of the risk averse students who were treated with multiple choice tests with NRET was lower than those who were treated with multiple choice tests with FS. However, differences in the average score of learning outcomes in the two treatments are not too large, this is because both scoring methods equally apply the penalty system on the wrong answer, it shows that risk averse students are more focused on a

penalty system as a risk in both scoring methods rather than on the opportunities that can be gained from NRET scoring method.

### Conclusion

The result of this study revealed that: (1) there are significant differences in chemistry learning outcomes among students who were given a multiple-choice test with NRET scoring method and FS scoring method, it means that there is a significant effect of multiple choice scoring method toward chemistry learning outcomes. NRET scoring method was found to be more effective and efficient to use in formative evaluation to enhance learning outcomes; (2) for risk seeking students, NRET scoring method gave a better result to maximize their learning outcomes than FS scoring method; (3) for risk averse students, FS provided higher learning outcomes than NRET; (4) effectiveness of multiple choice scoring methods on the formative evaluation were very dependent on the students' risk taking attitude, therefore both scoring methods can be used interchangeably in learning so that students with different risk taking attitude can still be facilitated by both.

### References

1. Ajayi, B.K. (1961). 'Effect of Two Scoring Methods on Multiple Choice Agricultural Science Test Scores,' *Review of European* 4 (1), 255-259.
2. Biria, Reza dan Ali Bahadoran B. (2015). 'Exploring the Role of Risk-Taking Propensity and Gender Differences in EFL students' Multiple-Choice Test Performance,' *Canadian Journal of Basic and Applied Sciences* 03 (05), 145-154.
3. Choppin, B. H. Correction for Guessing. In J. P. Keeves (Ed.). (1988). *Educational research, methodology, and measurement: an international handbook*, Pergamon Press, Oxford.
4. Coombs, C. H., Miholland, J. E., & Womer, F. B. (1956). 'The Assessment of Partial Knowledge,' *Educational and Psychological Measurement* 16, 13-37.
5. Crocker, Linda and J. Algina. (1986). *Introduction to Classical and Modern Test Theory*, Holt, Rinehart and Winston, Orlando.
6. Frary, Robert B. (1988). 'Formula Scoring of Multiple Choice Test (Correction for Guessing),' *Instructional Topics in Educational Measurement* 7 (2), 33-38.
7. Government Regulation No. 74 of 2008 on Teachers, Article 52, Paragraph 2
8. Gulo. W. (2005). *Research Methodology*, PT. Gramedia, Jakarta.
9. Howell, David C. (2007). *Statistical Methods for Psychology*. Sixth Edition, Thomson Wadsworth, Belmont, CA.
10. King, Bruce W., and Minium, M. Edward. (2003). *Statistical Reasoning in Psychology and Education*. Fourth Edition, NJ: John Wiley & Sons, River Street, Hoboken,
11. Lau, Paul. N. K., et al. (2011). "Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests," *Educational Technology & Society* 14 (4), 99-110.
12. Lau, Sie-Hoe, et al. (2014). 'Robustness of Number Right Elimination Testing (NRET) Scoring Method for Multiple-Choice Items in Computer Adaptive Assessment System (CAAS),' *Research and Practice in Technology Enhanced Learning* 9 (2), 283-300.
13. Lau, Sie-Hoe, et al. (2012). 'Web based Assessment With Number Right Elimination Testing (NRET) Scoring for Multiple-Choice item,' *The Asia-Pacific Education Researcher* 21(1), 107-116.

14. Middlecamp, C & E. Kean. (1985). Learning Guide Basic Chemistry, PT Gramedia, Jakarta.
15. Murray-Webster, Ruth dan David Hillson. (2007). Understanding and Managing Risk Attitude. Aldershot, UK: Gower.
16. Naga, Dali Santun. (2013). Scores Theory of Mental Measurements, PT. Nagarani Citrayasa, Jakarta.
17. Roediger, Henry L. & Elizabeth J. Marsh. (2005). 'The Positive and Negative Consequences of Multiple-Choice Testing,' *Journal of Experimental Psychology* 31 (5), 1155-1159.
18. Sanjaya, Wina. (2008). Curriculum and Learning: Theory and Practice Development Unit Level Curriculum, Kencana, Jakarta.
19. Suryabrata, Sumadi. (2000). Development of Psychological Measurement Tools, Penerbit ANDI, Yogyakarta.
20. Suryabrata, Sumadi. (2012). Educational Psychology, Raja Grafindo Persada, Jakarta.
21. Sudijono, Anas. (2007). Introduction of Educational Evaluation, PT Raja Grafindo Persada, Jakarta.
22. Sudjana, Nana. (2009). Teaching and Learning Outcomes Assessment, PT Remaja Rosdakarya, Bandung.
23. Weber, E.U., Blais, A.-R., & Betz, N. (2002). 'A domain-Specific Risk-Attitude Scale: Measuring Risk Perception and Risk Behavior,' *Journal of Behavioral Decision Making* 15, 263-290.
24. Witte, Roberts. (1985). Statistics. Second Edition, Donnelly & Sons Co., Pennsauken, NJ.