

THE EFFECT OF SCALE CATEGORY NUMBER ON ACCURACY OF PREDICTIVE VALIDITY OF MATHEMATICS SELF-EFFICACY SCALE

Iwan Setiadi

Madrasah Aliyah Al Wathoniyah 43 Jakarta
iwansetiadi_peg17s2@mahasiswa.unj.ac.id

Erdawaty Kamaruddin

Pascasarjana Universitas Negeri Jakarta
erda_kamaruddin@yahoo.com

Soeprijanto

Pascasarjana Universitas Negeri Jakarta
prianto.unj@gmail.com

ABSTRACT

The purpose of this study was to obtain empirical data about the effect of the number of scale categories on the accuracy of the predictive validity of the Mathematics Self-Efficacy Scale between those using a scale of three categories of answers and a scale of five categories of answers. This research uses a quantitative method of comparative study. The population in this study were all students of Madrasah Aliyah (Islamic High School) class XI IPA in North Jakarta in the academic year 2019/2020. Sampling was done by simple random sampling and there were 230 students from six public and private Madrasah Aliyah. The instruments used were tests and non-tests. The test instrument was in the form of a Mathematics learning achievement test and a non-test instrument in the form of a Mathematics Self-Efficacy Scale. The research hypothesis was tested using a t-test. The results showed that the average standard error of estimating the predictive validity of the Mathematics Self-Efficacy Scale using the five answer categories scale was lower than the scale of the three answer categories. The conclusion of the study is the predictive validity of the Mathematics Self-Efficacy Scale which uses a five-answer category scale that is more accurate in predicting mathematics learning outcomes.

Keywords: Mathematics self-efficacy, predictive validity, number of category scales, standard error of estimate

INTRODUCTION

Mathematics learning in schools is pursued through three aspects, namely: affective, cognitive, and psychomotor. However, in reality, Mathematics learning has been dominant in cognitive aspects. Though mathematics learning achievement is not only influenced by cognitive aspects but also influenced by other aspects, one of which is the affective aspect.

The importance of the affective aspect is because the mathematics learning outcomes of students who have a positive attitude in mathematics are higher than students who have a negative one (Hartati, 2015). That is the attitude towards the lesson, an important concern for improving learning outcomes. Concerning the ability to solve mathematical problems, affective factors that influence attitudes include mathematics self-efficacy (Güven & Ozum, 2013; Pimta, Tayruakham, & Nuangchalerm, 2009).

Self-efficacy is an individual's subjective perception of the belief in his ability to carry out tasks to achieve the desired results (Bandura, 1986) and the belief will be able to master the situation to give results as desired (Santrock, 2007). Self-efficacy has influences

cognitive learning outcomes (Firdaningsih, 2016; Ghufron& Suminta, 2013; Hemin Khezri azara, Masoud G. Lavasani, Ehsan Malahmadi, 2014; Liu & Koirala, 2009)and positively self-efficacy is associated with efforts to achieve student learning in answering test questions that are considered difficult (Salomon, 1984). The higher the student's self-efficacy, the higher the student's learning achievement (Susanti&Aulia, 2016).

In reality, there are still many students who have low mathematics self-efficacy. This is known from the actions, opinions, and negative attitudes of students towards Mathematics, for example, if the teacher gives an assignment then students are lazy to do it. If students are given a test, students assume that the test questions faced are different from the sample questions. It is also not uncommon for students to guess answers when taking a math test.

Self-efficacy can be used as a predictor of student learning success,(Andrew & Hons, 1998; Hemin Khezri azara, Masoud G. Lavasani, Ehsan Malahmadi, 2014; Nuo, Chao, Mcinerney, & Bai, 2018). Predictors mean that self-efficacy scores obtained by students can predict student success in certain subjects, including mathematics. For the predictor score to be used as a measure of self-efficacy, it is necessary to make an instrument that has validity tested, namely predictive validity. According to (Faleye, 2015), predictive validity refers to the extent to which a test can estimate a student's ability to carry out related activities in the future. Predictive validity is the ability of a test to predict what will happen in the future, for example, mathematics self-efficacy can predict mathematics learning outcomes for students of Madrasah Aliyah class XI IPA.

The instrument that can be used to measure mathematics self-efficacy is a questionnaire that uses a Likert Scale. Likert Scale is a scale used to measure the attributes of attitudes and opinions through a questionnaire (Naga, 2013). In the questionnaire, positive and negative statements are mixed randomly. This is intended to prevent the respondent from agreeing to the questionnaire without reading it again. With the incorporation of positive and negative statements in the questionnaire, respondents are urged to read each statement carefully before the respondent answers. So getting the measurement score as the real thing. Instruments on the Likert Scale are designed with a question or statement that responds with answers to agree or disagree, like or dislike (Adelson & Mccoach, 2010) or accordingly or not (Azwar, 2016) in the range of one to five answer categories (Likert, 1932).

The current self-efficacy instruments, in terms of the number of answer category scales, vary in number. There are self-efficacy studies that use a Likert Scale with a total of four categories (Kalaycioğlu, 2017), five-category scale (Nuo et al., 2018; Ozel, Caglak, & Erdogan, 2013) and some use a scale of six categories (Koyuncu, Guzeller, & Akyuz, 2017; Usher & Pajares, 2009), some even use a scale of ten categories (Zarch & Kadivar, 2006). Of course, this can make doubts that there are some appropriate category scales for measuring students' mathematics self-efficacy. For that self-efficacy scale instrument used in order to predict student learning outcomes it is necessary to test its validity through predictive validity and also reliability testing.

Several studies have been conducted on the differences in the number of scale categories in attitude measurement. Among other studies conducted by(Adelson & Mccoach, 2010; Chang, 2014; Lozano, García-cueto, & Muñiz, 2008; Preston & Colman, 2000(Kim, 2010). These studies discuss the effect of the number of category scales on

validity and reliability. In general, the findings of the study are that the increasing number of categories of scale increases the coefficient of validity and reliability. But the increase in the coefficient of validity and reliability is only on the number of certain scale categories. The number of scale categories is not guaranteed to be high also the coefficient of validity and reliability.

The variety of scale categories in some of these studies, so far there have been no studies that examine the effect of the number of scale categories on the accuracy of predictive validity. Existing research is comparing the level of validity and reliability based on the number of scale categories. In contrast to previous studies, this study examines the effect of the number of scale categories on the accuracy of predictive validity by looking at the average standard error of estimate. The number of scale categories in this study is three and five. Determination of the scale of three categories because it is more similar to the Likert scale, which has a middle category in measuring attitudes (Saragih, 2017). In addition, the middle category is a neutral, more reliable and preferred response from respondents (Adelson & McCoach, 2010; Cronbach, 1950). Whereas the scale of the five categories of answers is original from the number of categories on the Likert Scale itself in measuring attitudes (Likert, 1932).

The choice of the number of scale categories on the attitude scale affects the level of instrument reliability, namely the score on the attitude instrument with the number of answer category scales, tend to have greater variance, then the reliability becomes higher (Mueller, 1986) in (Hadi & Pinang, 2013). If the number of alternative answer categories increases, reliability and validity increases (Lozano et al., 2008). With regard to validity, that scales with relatively few response categories tend to produce scores with relatively small variances so, limiting the magnitude of the correlation (Preston & Colman, 2000). Variance also affects the level of validity, because the correlation itself is influenced by the value of variance and covariance (Naga, 2013). As the following formula :

$$\sigma_{xy} = \rho_{xy} \cdot \sigma_x \cdot \sigma_y$$

Note :

- σ_{xy} = covariance between x and y
- ρ_{xy} = correlation
- σ_x = standard deviation x
- σ_y = standard deviation y

The validity coefficient is based on the correlation coefficient between the predictor score and the criterion score. Correlation is related to linear regression. One of the objectives of regression analysis is to predict. In the regression equation needs to be tested for accuracy in predicting. Its accuracy is analyzed through the standard error of estimate (SEE). As said by Azwar (2018) that to determine predictive validity in addition to using correlations it is necessary to be accompanied by data on the magnitude of the standard error of the estimate. The relationship between the standard error in estimate and the validity coefficient is the same as the relationship between the standard error in measurement with the reliability coefficient. The higher the correlation, the closer the relationship between the predictor and the criteria (Klapproth & Schaltz, 2014) so that the smallest SEE value is the most accurate predictive validity.

The magnitude of the SEE shows a measure of the standard distance between the estimated value of y on the regression line and the actual y value obtained from the sample (Morrison, 2016). The smaller the SEE value, the higher the accuracy of the estimated equation produced to explain the value of the variable \hat{y} (Pratomo, 2015). SEE is directly related to the magnitude of the correlation between x and y . If the resulting correlation level approaches $+1.00$ or -1.00 then the values converge close to the regression line, and the standard error estimate shows a small value, likewise, if the correlation is close to zero, then the values will spread away from the regression line (Morrison, 2016). That causes the ability of the regression equation to predict to be reduced so that the SEE becomes large. The coefficient of determination r^2 , can be obtained if the correlation coefficient between x and y is raised to two. The coefficient of determination can determine how much difference (variability) in y that can be predicted (predicted variability) with x . r^2 measures the level of predictable variability on the y score, called $SS_{regression}$, so the unpredictable part, $1 - r^2$ is called $SS_{residuals}$. Thus the Standard error of estimate (SEE) is calculated using the following formula (Morrison, 2016).

$$s_{yx} = \sqrt{\frac{SS_{residual}}{df}} = \sqrt{\frac{\sum(y - \hat{y})^2}{n-2}}$$

Note:

s_{yx} = standard error of estimate

y = score of observational data

\hat{y} = predictive value

n = size of data pair

df = degree of freedom

Based on this, by knowing the size of the value of the SEE can be known the accuracy of predictive validity. To compare the average SEE predictive validity of the Mathematical Self-Efficacy Scale between those using a scale of three categories of answers and a scale of five categories of answers it is necessary to do research. So that the Mathematical Self-Efficacy Scale instrument can be used to predict mathematics learning outcomes by considering the number of scale categories.

METHOD

This research method is quantitative of comparative study. The population is all students of class XI IPA Madrasah Aliyah in North Jakarta. For the selection of school samples and student samples carried out by random sampling. The sample came from six Madrasah Aliyah, they are ; MAN 5, MAN 21, MA AL WATHONIYAH 43, MA YAPIS, MA AL KHAIRIYAH, and MA AL JIHAD. The total sample was 230 students who were divided into two groups, namely: 115 students using a scale of three categories and 115 students using a scale of five categories.

Validity The contents of the research instrument have been tested materially by experts and the construct validity was tested using the Confirmatory Factor Analysis (CFA) approach with the help of the LISREL 8.72 program. Testing the instrument on 400 students of Madrasah Aliyah Class XI IPA obtained three and five scale instrument

reliability are 0.82 and 0.74. The test instrument was in the form of Matrix subject matter in class XI, while the test scores were in the form of a Mathematics Self-efficacy Scale with some categories three and five. For the scale of three categories of choices, consisting of favorable statements, namely: 1 (very inappropriate) to 3 (very appropriate). In contrast to unfavorable statements starting from 1 (very appropriate), to 3 (very inappropriate). As for the scale of the five-categories, consisting of favourable statements are 1 (highly inappropriate), to 5 (very appropriate) Conversely for unfavorable statements are starting from 1 (very appropriate), to 5 (very inappropriate). Research data were analyzed using independent sample *t*-tests. Research design as Table. 1 of the following

Tabel 1. Research Design

Resampling Order	3-Scale Category	5-Scale Category
	S_{yx}	S_{yx}
1	$S_{1\ yx}$	$S_{1\ yx}$
2	$S_{2\ yx}$	$S_{2\ yx}$
3	$S_{3\ yx}$	$S_{3\ yx}$
.	.	.
.	.	.
.	.	.
25	$S_{25\ yx}$	$S_{25\ yx}$
Average	μ_{s3k}	μ_{s5k}

Note:

s_{yx} = standard error of estimate (SEE)

μ_{s3k} = average SEE of 3-scale category answers

μ_{s5k} = average sEE of 5-scale category answers

The independent variable in this study is the number of answer category scales on the Mathematics Self-efficacy Scale, which is a scale of three and a scale of five categories. While the dependent variable is the average SEE predictive validity.

The research procedure was as follows: (1) A total of 230 students were made into two groups of respondents to fill in the same the Mathematics Self Efficacy Scale. 115 students fill the five-category scale and 115 other students fill the three-category scale, (2) Furthermore, in the period of one to two weeks after students fill the Mathematics Self-Efficacy Scale then students work on the learning achievement test, which is the daily assessment of the subject matrix , (3) Analyzing correlation and simple regression between Mathematics Self-Efficacy Scale and Mathematics learning outcomes of Matrix subjects to find out the equation of the regression model, (4) Performing resampling using Excel program in each group with 25 times of resampling and each resampling taken 30 samples by way of return. (5) Determine the SEE value at each time using the Excel resampling program, (6) Determine the average of the SEE value for each group, (7) Statistical test using the *t*-test, which compares the average SEE value of the two groups.

RESULT

Data description from 25 times the resampling group predictive validity Mathematics Self-efficacy Scale using a five-category scale obtained an average SEE of 2.852 with a variance of 0.094. Read more like Table.2

To test inferential statistical hypotheses it is necessary to conduct prerequisite tests of data analysis, namely tests of normality and homogeneity of data. The results obtained are as follows: Test for normality using Liliefors, the results are $L_0 = 0,136$ meanwhile $L(t) = 0,173$ at the level $\alpha = 0,05$ with a total of 25 data. This means that $L_0 < L(t)$, then accept H_0 or in other words the data comes from populations that are normally distributed.

Table 2. Data Resampling Results Mathematics Self-efficacy Scale of Five Categories

Resampling Order	SEE	Resampling Order	SEE	Resampling Order	SEE
1	2.294	11	3.190	21	2.818
2	2.714	12	2.417	22	2.927
3	3.111	13	2.838	23	3.231
4	2.990	14	2.897	24	2.918
5	3.034	15	2.966	25	2.961
6	1.972	16	3.294		
7	2.913	17	2.869		
8	2.847	18	2.505		
9	2.732	19	3.280		
10	2.664	20	2.914		

For the predictive validity group Mathematics Self-efficacy Scale which uses a scale of three categories, which is obtained an average SEE of 3.386 with a variance of 0.062. Read more like Table.3.

Table 3. Data Results of Resampling Mathematics Self-efficacy Scale of Three Categories

Resampling Order	SEE	Resampling Order	SEE	Resampling Order	SEE
1	3.652	11	3.003	21	3.232
2	3.544	12	3.464	22	3.686
3	3.312	13	3.433	23	3.387
4	3.585	14	3.148	24	3.574
5	3.477	15	3.139	25	3.027
6	3.164	16	2.688		
7	3.574	17	3.208		
8	3.522	18	3.539		
9	3.580	19	3.637		
10	3.577	20	3.484		

As for the calculation of normality test data the average value of the SEE Mathematics Self-Efficacy Scale of three answer categories was obtained $L_0 = 0,144$

meanwhile $L_t = 0,173$ at the level $\alpha = 0,05$ with a total of 25 data. This means that $L_0 < L_t$, then accept H_0 or in other words the data comes from populations that are normally distributed.

Next is the data homogeneity test using the F-test. Criteria: accept H_0 if $F_c < F_t$, means that both groups have the same or homogeneous variance. Calculation results are obtained $F_c = 1,520$ and $F_t = 1,980$ at the level $\alpha = 0,05$ and $df = 24, df = 24$, then $F_c < F_t$ or $1,520 < 1,980$, means both groups are homogeneous.

The comparison of the average SEE data comparison for a scale of five categories and three categories as shown in Figure 1 below.

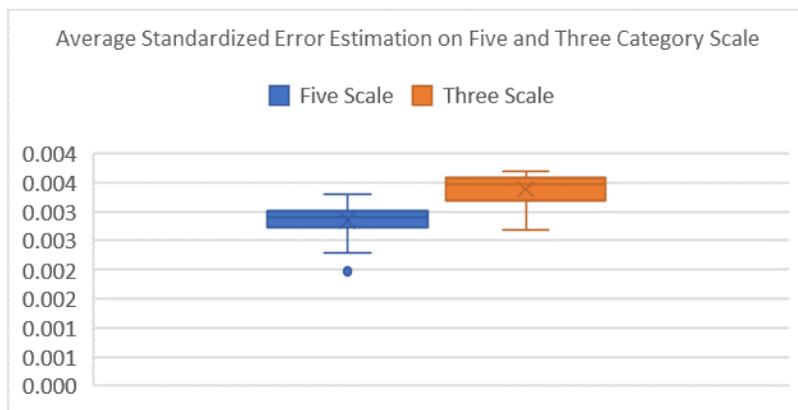


Figure 1. Boxplot of Average *Standard Error of Estimate (SEE)*

In Figure 1. it can be seen that for a scale of five categories, the shape of the average SEE data distribution is more normal than a scale of three categories. On a scale of three, the average SEE data tends to collect at large values. On a scale of five categories data is more homogeneous compared to a scale of three categories.

Next is testing the statistical hypothesis with the t-test. The criteria is to accept H_0 if $t_c < t_t$. Based on the results of the t-test or the two difference test the average for the free sample is obtained $t_c = 6,743$ meanwhile t_t with $df = 48$ at $\alpha = 0,05$ or $t_t (0,05; 48)$ of 1.677. Then $t_c > t_t$ which means reject H_0 , the average standard error of estimating the predictive validity of the Mathematics Self-Efficacy Scale which uses a five-answer scale is lower than the scale of three answer categories.

DISCUSSION

The results showed that there were differences in the average standard error of estimate of the predictive validity of the Mathematical Self-Efficacy Scale between those using a scale of three and a scale of five categories of answers. The scale of the five categories of answers has average SEE lower than the scale of the three categories of answers. This is because the scale of the five-categories has a higher variance than the scale of the three-categories. High variance influences validity or correlation. This study is in line with the results of the study of (Lozano et al., 2008), which compares the number of answer category scales between two to nine category scales. The results show that the more scale the number of answer categories, the more reliability and validity, besides it is explained that the optimal number of scales is between four and seven-categories while the scale category is less than four, the reliability and validity decreases. (Preston & Colman,

2000)Research also supports this research by finding that reliability and validity for a scale of two, three, and four categories have relatively poor performance compared to a scale with more answer categories, up to about seven.

The research findings of (Gupta, 2014) regarding students' attitudes toward their studies are different from the findings of this study. They stated that there was no significant effect between the number of scales of two, three or five categories on the Likert scale on reliability and validity. When viewed from the sample used in the study, the sample is not homogeneous. The number of 510 students came from different study programs, namely: Physical Education, Art Education, Special Education, Islamic Studies, and Psychology. In making affective instruments, in addition to considering the number of category scales, it should also pay attention to the characteristics of the sample in order to have high reliability (Hadi & Pinang, 2013). The reliability and validity coefficient depends on the size of the variance. Variance influences correlation and regression, and the correlation itself is none other than validity. In accordance with the results of (Retnawati, 2015), that the higher the measurement accuracy, the smaller the standard measurement error, the higher the accuracy of predictive validity the smaller the SEE.

CONCLUSION

The results of this study are the average SEE Mathematics Self-efficacy Scale which uses a scale of five categories lower than the scale of three categories. Thus the predictive validity of the most accurate Mathematics Self-efficacy Scale is that of using a five-category answer scale.

REFERENCES

- Adelson, J. L., & McCoach, D. B. (2010). Educational and Psychological Measurement. <https://doi.org/10.1177/0013164410366694>
- Andrew, S., & Hons, B. (1998). Self-efficacy as a predictor of academic performance in science, 596–603.
- Azwar, Saifuddin. “Sikap Manusia Teori dan Pengukurannya.” Yogyakarta: Pustaka Pelajar. 2016
- Azwar, Saifuddin. “Reliabilitas dan Validitas.” Yogyakarta: Pustaka Pelajar. 2018.
- Bandura, A. (n.d.). The Explanatory And Predictive Scope Of Self-Efficacy Theory, 4(i), 359–373.
- Chang, L. (2014). Applied Psychological Measurement A Psychometric Evaluation of 4-Point and and Validity. <https://doi.org/10.1177/014662169401800302>
- Cronbach, L. J. (1950). Educational and Psychological Measurement. <https://doi.org/10.1177/001316445001000101>
- Faleye, B. A. (2015). Predictive Validity of Students ' Entry Qualifications into Mathematics Programme in Nigeria ' s Osun and Oyo States ' Colleges of Education, 4(4), 209–217. <https://doi.org/10.15640/jehd.v4n4a25>
- Firdaningsih. (2016). Efikasi Diri Dan Motivasi Berprestasi Dalam Meningkatkan Hasil

- Belajar Sejarah, *J*, 47–60. <https://doi.org/10.22236/JPPP>
- Ghufron, M. N., Suminta, R. R., & Psikologi, P. S. (2013). Efikasi Diri dan Hasil Belajar Matematika :, *21*(1), 20–30.
- Gupta, S. (2014). The Effect of Number of Alternatives on Validity and Reliability in Likert Scale ISSN 2319-9725, (February). <https://doi.org/10.13140/2.1.2237.2803>
- Güven, B., & Özüm, B. (2013). Factors in influencing mathematical problem-solving achievement of seventh grade Turkish students. *Learning and Individual Differences*, *23*, 131–137. <https://doi.org/10.1016/j.lindif.2012.10.003>
- Hadi, S. M., & Pinang, P. (2013). The Comparison Of Number Response Categories Towards Reliability Of Mathematics Disposition, *4*(2), 105–117.
- Hartati, L. (2015). Pengaruh Gaya Belajar Dan Sikap Siswa Pada Pelajaran Matematika Terhadap Hasil Belajar Matematika. *Formatif*, *3*(3), 224–235.
- Hemin Khezri azara, Masoud G. Lavasani, Ehsan Malahmadi, J. A. (2014). The role of self- efficacy , task value , and achievement goals in predicting learning approaches and mathematics achievement. *Procedia - Social and Behavioral Sciences*, *5*(December 2010), 942–947. <https://doi.org/10.1016/j.sbspro.2010.07.214>
- Kalaycıoğlu, D. B. (2017). The Big Fish-Little Pond Effect on Affective Factors Based on PISA 2012 Mathematics Achievement PISA 2012 Matematik Başarısına Dayalı Duyuşsal Faktörlerde Büyük Balık - Küçük Göl Etkisi, *8*(1), 1–14. <https://doi.org/10.21031/epod.297686>
- Klapproth, F., & Schaltz, P. (2014). The Validity Of Predictors Of Academic And Vocational-Training Achievement : A Review Of The Literature, 153–166.
- Koyuncu, I., Guzeller, C. O., & Akyuz, D. (2017). The development of a self-efficacy scale for mathematical modeling competencies, *4*(1), 19–36. <https://doi.org/10.21449/ijate.256552>
- Likert, R. (1932). Likert, Rensis. “A Technique For The Measurement of Attitudes.” Archives Of Psychology, New York. 1932.
- Liu, X., & Koirala, H. (2009). The Effect of Mathematics Self-Efficacy on Mathematics Achievement of High School Students.
- Lozano, L. M., García-cueto, E., & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales, *4*(1996), 73–79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Morissan. “Statistik Sosial.” Jakarta: Kencana. 2016.
- Naga, Dali S. “Teori Skor Pada Pengukuran Mental”. Jakarta: Nagarani Citrayasa. 2013.
- Nuo, C., Chao, G., Mcinerney, D. M., & Bai, B. (2018). Self-efficacy and Self-concept as Predictors of Language Learning Achievements in an Asian Bilingual Context. *The Asia-Pacific Education Researcher*. <https://doi.org/10.1007/s40299-018-0420-3>
- Ozel, M., Caglak, S., & Erdogan, M. (2013). Are affective factors a good predictor of

science achievement? Examining the role of affective factors based on PISA 2006. *Learning and Individual Differences*, 24, 73–82.
<https://doi.org/10.1016/j.lindif.2012.09.006>

- Pimta, S., Tayruakham, S., & Nuangchalerm, P. (2009). Factors Influencing Mathematic Problem-Solving Ability of Sixth Grade Students Chumchonbanchomsa-Ard School , Moeiwadi District , Roi-et Province 45250 Thailand Department of Research and Development Education , Faculty of Education , Department of Curriculum and Instruction , Faculty of Education , Mahasarakham University , Mahasarakham 44000 Thailand, 5(4), 381–385.
- Pratomo, Dedi S., dan Astuti, Erna Z. “Analisis Regresi Dan Korelasi Antara Pengunjung Dan Pembeli Terhadap Nominal Pembelian Di Indomaret Kedungmundu Semarang Dengan Metode Kuadrat Terkecil.” 2015. *Jurnal Statistika*.
http://eprints.dinus.ac.id/16877/1/jurnal_15951.pdf (diakses 25 April 2019)
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales : reliability , validity , discriminating power , and respondent preferences, 104, 1–15.
- Salomon, G. (1984). Television Is " Easy " and Print Is " Tough ": The Differential Investment of Mental Effort in Learning as a Function of Perceptions and Attributions, 76(4), 647–658.
- Susanti, T., Aula, U., Biologi, P. P., & Ilmu, F. (2016). Mata Pelajaran Ilmu Pengetahuan Alam Terpadu, 1(1), 34–41.
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics : A validation study, 34, 89–101. <https://doi.org/10.1016/j.cedpsych.2008.09.002>
- Zarch, M. K., & Kadivar, P. (2006). The Role of Mathematics self-efficacy and Mathematics ability in the structural model of Mathematics performance, 2006, 242–249.