



Empirical Test Of Content Validity And Construct Validity Of Mathematical Self-Efficacy Scale

Iwan Setiadi

Madrasah Aliyah Al Wathoniyah 43 Jakarta Utara, Indonesia

Email: iwansetiadi43@gmail.com

Website: <http://jurnal.uin-antasari.ac.id/index.php/jtijk>

Received: 3 January 2021; Accepted: 5 October 2021; Published: 31 December 2021

ABSTRACT

The purpose of the study is to examine the instrument content validity and it's construct validity of the mathematics self-efficacy scale based on valid dimensions. This research is part of the research on the effect of the number of scale categories on the accuracy of the prediction validity of the mathematics self-efficacy scale. The research instrument is a Likert scale questionnaire with five answer categories. Analysis of content validity using coefficient V by Aiken while construct validity uses the Confirmatory Factor Analysis method with the help of the LISREL 8.72 program. The results showed that (1) 28 of 40 items of the instrument content validity is valid in three dimensions : level, strength, and generality. (2) While 22 items out of 28 items of the instrument construct validity is proven valid as well. In conclusion, the mathematics self-efficacy scale instrument has been empirically tested valid for its content and construct validity. **Key Words:** mathematics self-efficacy scale, content validity, construct validity, self-efficacy.

INTRODUCTION

So far, mathematic learning process at school has a purpose to increase the understanding of mathematical concept (Purnama, Wijaya, Dewi, & Zulfah, 2020). However, mathematic learning achievement is not influenced by the cognitive aspect only but also influenced by the other aspects, one of them is the affective aspect. The affective aspect is giving a contribution and a success in Mathematic learning (Alam, 2018). Beside of that, mathematic learning result of students who have positive behavior toward mathematic subject is higher than who have negative behavior toward mathematic subject (Hartati, 2013). It means, the behavior towards a subject, especially mathematic subject is being the important attention to increase learning result.

One of the affective factor that affects the behavior to mathematic learning result is: mathematical self-efficacy (Guyen & Cabakcor, 2013). Self-efficacy is a value or individual subjective perception to the

belief of self ability in organizing and deciding a needed behavior to reach a wanted performance (Bandura, 1986). There are 4 sources of self-efficacy, which are: Enactive Mastery Experience, Vicarious Experience, Verbal Persuasion, and Physiological and Affective States (Bandura, 1997).

Self-efficacy is a belief that someone can control a situation and give an appropriate result as what the want is (Santrock, 2007). A student who has low self-efficacy might do not want to try anymore to do a mathematic assignment because he / she feels unconfident that learning can help him / her to do an easy to difficult question. Whereas, in mathematic, there is a pre-condition material for students to be able to follow the next material. Mathematic is a knowledge about the organized structure from an undefined element to the defined element, to axiom or postulate then to proposition or theorem (Russeffendi, in Kusri, et al. 2014). While, student who has high self-efficacy believes

that he / she can do the assignments. Therefore, in academic purpose, students who have high self-efficacy academic believe that they can do the academic assignment. For example, mathematical self-efficacy means that students believe they can success in Mathematic subject and can do the mathematic assignment even though it is difficult.

Mathematic self-efficacy is a belief or individual perception that relate to the ability in mathematic. The previous researches state that self-efficacy has a correlation with learning result (Firdaningsih, 2016; Ghufron & Suminta, 2013) and affects to the cognitive learning result (Rosyida, Utaya, & Budijanto, 2016).

The reality in the field is, there are still many students who have low mathematical self-efficacy (Arifin & Asdar, 2015). It can be known from the students' negative action, opinion, and behavior towards mathematic subject (Sunaryo, 2017). When students face a daily assessment, then students who have high mathematic self-efficacy should success in answering mathematic question so that the mathematic learning result is high.

Self-efficacy can be a predictor of students learning success, as the result of research (Andrew, 1998; Chao, McInerney, & Bai, 2019). A predictor means that with seeing the student's self-efficacy score, it can predict students' success in learning a certain subjects. In order to the predictor score can be the measuring instrument of self-efficacy, then it needs to be made an measuring instrument that the validity is tested that known as predictive validity.

The meaning of instrument is mathematic self-efficacy scale. This scale should be really can to predict students' success when answering the test in the daily assessment.

The scale is a set of questions that arranged to reveal a certain attributes through a respond towards the attribute question (Azwar, 2018). While Naga (2013) said that the scale is a rule of numbering to the attribute. A used number is a real number

that has several characteristics. The scale has a scale level, which are: First, the nominal scale that only has one information which is a difference. Second, the ordinal scale that gives information about a difference and level. Third, the interval scale that has information about three things, which are; differentiate, level, and a same distance. Fourth, the ratio scale that has most informations, which are; differentiate, level, a same distance, and has an absolute zero.

The mathematic self-efficacy made should has predictive validity. Azwar (2018a) explained about the predictive validity, which is: if a test that is arranged to predicts the performance in the future, then the test should has a predictive function that validated by a relevant score criteria.

An instrument can called as has a quality if it fulfils a valid criteria (Istiqamah, 2019). The mathematic self-efficacy scale instrument that used should has some conditions, which are valid and reliable (Mardapi, 2012). A test validity should be determined to know the quality of the test in the relation with measure a thing that should be measured. According to Naga (2013), a measurement validity is a compatibility between the measurement and measuring target which is with what the thing that want to be measured is. The definition of validity is always be related with an empirical research and the proof depends on the type of validity that is used. Therefore, the validity can be used in checking directly how far a test tool is working.

Azwar (2018a) explained that content validity is a validity that estimated through feasibility test or the relevance of content test with the measuring purpose and rationally analyzed by a competent panel (expert judgment). The indicator of an instrument called as valid if there is no other revise by the expert (Yusup, 2018). Because content validity uses some experts in a certain field, so the assessment is subjective. Nevertheless, the assessment agreement from the experts still can be estimated empirically.

Allen & Yen in Azwar (2018a) explained that construct validity is a validity that shows how far the test result can reveal a theoretical construct trait that wants to be measured. The construct validity proves is a measurement result that obtained through the test items highly correlated with the theoretical construct. Is the obtained score support the theoretical concept that is wanted by the first measurement purpose.

Based on that, the test empirically that relates with the mathematic self-efficacy scale construct validity is important to be done. This has a purpose to know is the quality of the measurement tool has appropriate with the theoretical construct (Widodo, 2006).

Remembering about how important the benefit of self-efficacy scale instrument is so that can be predict mathematic learning result so this instrument has done a test. There is also the test that performed is test empirically the mathematical self-efficacy scale instrument, based on the content and construct validity.

RESEARCH METHOD

The research method is quantitative. The training instrument test sample is taken from two high schools (MAN) class XI IPA, each of it is at Jakarta and Bekasi. The amount of sample is 202 students, which are 68 boys and 134 girls. The mathematic self-efficacy scale instrument is developed based on self-efficacy dimension (Bandura, 1997), which are: (1) Level, (2) Strength, and (3) Generality. Each dimensions are divided into several indicators as in Table 1.

Table 1. Dimension and Indicator of Mathematic Self-Efficacy

Dimension	Indicator
1. Level	1.1. The ability to finish assignments
	1.2. The time in finishing an assignments
2. Strength	2.1. Self-assessment to the ability
	2.2. Others-assessment to the ability

	2.3. The Focus in Finishing an Assignments
3. Generality	3.1. Mastery of a Certain Assignment
	3.2. The Ability of Finishing Assignment in a Certain Condition

Then, the mathematic self-efficacy scale is made using Likert scale with five answer categories, which are Very Inappropriate (*Sangat Tidak Sesuai – STS*), Inappropriate (*Tidak Sesuai – TS*), Between Appropriate and Not (*Antara Sesuai dan Tidak – N*), Appropriate (*Sesuai – S*), and Very Appropriate (*Sangat Sesuai – SS*). According to Azwar (2016), ideally, a scale of Likert's model attitude should consist of favorable statement and unfavorable statement which are relatively balanced.

The number of self-efficacy statement item for the statement is 40 items. This is appropriate with the statement (Alwi, 2012), that it is recommended to doubling the number of item that will be used as the research instrument. The highest score for the mathematic self-efficacy scale uses Likert's Scale five answer categories is $5 \times 40 = 200$ and the lowest score is $1 \times 40 = 40$. The mathematic self-efficacy scale instrument which will used in this research are as much 20 items.

The content validity

The content validity test in this research uses three experts, which are Psychology, Assessment, and Instrument Development expert. The content validity item is proved with using Aiken's index validity (V Aiken). The coefficient of Aiken's validity has a range between -1 until 1. The higher V Aiken coefficient, the better the statement item (Naga, 2013). The V Aiken validity coefficient (Aiken, 1985) is gained with applying the formula:

$$V = \frac{\sum(r_i - l_0)}{[n(c-1)]} \quad (1)$$

Description:

r = the number that is given by the evaluator
 l_0 = the number of the lowest validity score
 c = the number of the highest validity score

n = the number of expert & practitioner who are doing the scoring

i = integers from 1, 2, 3 until n

The construct validity

The construct validity observes a construct attribute in a measuring instrument. According to Mardapi (2012), the construct validity refers to how far the test measures the concept from a theory, which is what the thing that be a base in test arrangement is. The construct validity estimation is not stated in coefficient form. The approach that is widely used in the construct validity test is multitrait-multimethod approach and factor analysis approach (Azwar, 2018a).

In this research, the construct validity is tested with using factor analysis approach, which is Confirmatory Factor Analysis (CFA). The data is analyzed using LISREL 8.72 program. LISREL is the most informative program in presenting the statistic results until the good and the bad in the research model can easily known (Ghozali & Fuad, 2014).

The research method can be said as good if having a good *fit* model. A *fit* model is the accuracy of assessing a research method. The indicator for assessing the *fit* model, inter alia are: *Chi-Square* and probability ($p \geq 0,05$); Goodness of Fit Indices ($GFI > 0,9$); Adjusted Goodness of *Fit* Index ($AGFI \geq 0,9$); Root Mean Square Error of Approximation ($RMSEA < 0,05$) means the *fit* model (Ghozali & Fuad, 2014). If the score of RMSEA is around 0,08, it means the model has the reasonable estimated error, while RMSEA is around between 0,08 until 0,1, it means the model has enough *fit*, and if RMSEA is bigger than 0,1, it means the model is very bad (MacCallum at al. in Ghozali & Fuad, 2014).

If the chi-square result is not significance ($p > 0.05$), then the hypothesis is nil “not rejected” (Febriana, 2015). It means the unidimensionality can be accepted if the item or the instrument sub test is only measuring one factor. While. If the Chi-Square score is significance ($p < 0.05$),

it means that the item is measuring more than one factor or character multidimensional. Then, the measurement model is needed to be modified. Beside of that, the factor load is significance if ($t > 1,96$) and positive.

RESULTS AND DISCUSSION

The Content Validity

The V Aiken coefficient from the mathematic self-efficacy scale with five answer categories and the number of 40 statement items is 0,67 with the probability 0,009. It refers to the criteria of V Aiken coefficient minimum score in the number of five categories and the number of 25 statement items (Aiken, 1985). The more experts who study the instrument, then the better of the instrument quality in the content validity aspect (Bashoor & Supahar, 2018). The average score of V Aiken coefficient is obtained in the result as this:

Table 2. The Average Score of V Aiken’s Mathematic Self-Efficacy Scale

Mathematic Self-Efficacy Indicator	The Average of V Aiken
The Ability to Finish the Assignment	0,85
The Time in Finishing the Assignment	0,75
Self-Assessment to the Ability	0,63
Others-Assessment to the Ability	0,75
The Focus in Finishing the Assignment	0,68
The Mastery of a Certain Assignment	0,68
The Ability to Finish the Assignment in a Certain Condition	

In the indicator of the ability to finish the assignment (KD), the average of V Aiken coefficient is high, which is 0,85. Nevertheless, from seven statement items (item 1st, 2nd, 4th, 5th, 6th, 31st, and 39th), there is one item fall, which the item number 39. The 39th item read *Saya belajar sebelum ulangan matematika* (I learn before mathematic test). The average of V Aiken coefficient in the indicator of time of finishing the assignment (*waktu penyelesaian tugas – WT*) is 0,75. Consist of 5 statement items, which are the item 3rd, 11th, 23rd, 29th, and 38th. Those five items is called valid. The self-assessment to the

ability indicator (*penilaian diri terhadap kemampuan – PD*) consist of six item statements. From the six items, there are only two that called void, which are; the item number 7, read I believe that I can do the mathematics question with my own way (*saya yakin dapat mengerjakan soal matematika dengan cara saya sendiri*), and the item number 25, which is: I have a problem to do the entire mathematics assignment (*saya memiliki kendala mengerjakan seluruh tugas matematika*). So, the average is very little, which is 0,60 as what is seen in the table 2.

The indicator of others-assessment to the ability (*penilaian orang lain terhadap kemampuan – PO*) has the V Aiken coefficient average at 0,63. From six statement items (item 15th, 19th, 20th, 21st, 22nd, and 34th), the three item that is not valid are the 15th, 20th, and 34th items. The 34th item read as My friends are mocking me when I do wrong in the assignment so that makes me lazy to learn (*teman-teman mengejek ketika saya salah mengerjakan tugas sehingga membuat saya malas belajar*). Then, the indicator of the focus in finishing the assignment (*fokus dalam menyelesaikan tugas – FT*) has V Aiken coefficient average at 0,75. From the five statement items (item 12th, 17th, 24th, 35th, and 40th) are all valid. While the indicator of the mastery of a certain assignment (*penguasaan tugas tertentu – PT*) has V Aiken average at 0,68. The item is item 13th, 14th, 27th, 28th, and 32nd. The 27th item that is not valid, read as I am lazy to do a difficult questions (*saya malas mengerjakan soal-soal yang sulit*). The last indicator is the ability to finish the assignment in a certain condition (*kemampuan menyelesaikan tugas dalam kondisi tertentu – KT*) has V Aiken coefficient average at 0,68. From six items (item 16th, 18th, 26th, 30th, 33rd, and 37th) there are two items that is not valid which are the 26th and 30th item. The 30th item read as a good mood affects me to finish the mathematics assignment on time (*suasana hati yang menyenangkan mempengaruhi*

saya menyelesaikan tugas matematika tepat waktu). The full result is as in the table 3.

Table 3. The Result of Content Validity Test

No Item	Expert 1	Expert 2	Expert 3	V Aiken Coefficient
1	5	5	5	1,00
2	5	5	4	0,92
3	4	4	4	1,00
4	5	5	5	1,00
5	5	5	3	0,83
6	4	5	4	0,83
7	5	5	4	0,92
8	2	3	2	0,33
9	2	4	3	0,50
10	2	4	2	0,42
11	4	5	4	0,83
12	4	5	4	0,83
No Item	Expert 1	Expert 2	Expert 3	V Aiken Coefficient
13	4	3	4	0,67
14	2	5	4	0,67
15	4	4	2	0,58
16	4	4	4	0,75
17	4	4	4	0,75
18	4	4	4	0,75
19	4	3	4	0,67
20	4	3	2	0,50
21	4	4	4	0,75
22	4	4	4	0,75
23	4	4	4	0,75
24	4	4	4	0,75
25	5	4	4	0,83
26	3	3	4	0,58
27	4	4	2	0,58
28	4	4	4	0,75
29	2	4	4	0,75
30	2	3	4	0,50
31	4	5	3	0,75
32	4	4	4	0,75
33	4	4	4	0,75
34	4	3	2	0,50
35	4	4	4	0,75
36	2	4	4	0,58
37	4	4	4	0,75
38	2	3	3	0,42
39	2	5	3	0,58
40	4	3	4	0,67

Based on the table 3, from 40 statement items, there are 12 statement items that is not appropriate to the V Aiken coefficient criteria $\geq 0,67$ so that is called as invalid. Those statement item is not opted-in in the construct validity.

The Construct Validity

The test of construct validity is done with sorting and grouping appropriate to the indicator. The construct validity test uses Confirmatory Factor Analysis (CFA) with

the help of LISREL 8.72 program. The result as seen in the Picture 1, that concluded in Table 4 below.

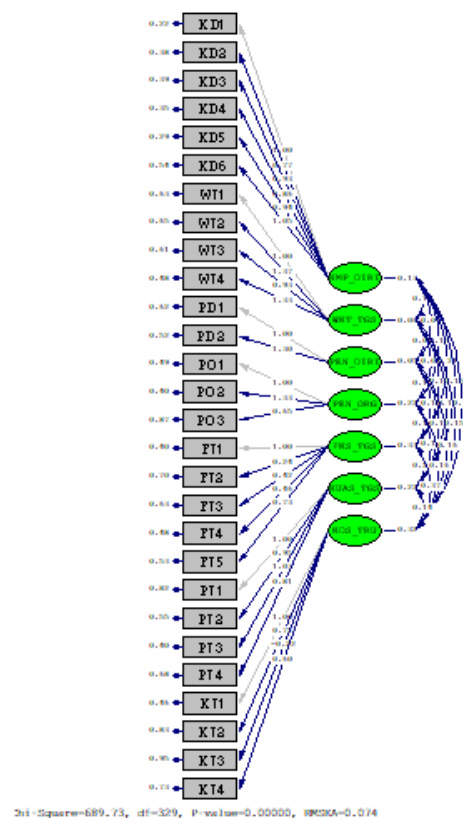
Table 4. The Result of Construct Validity Test

Indicator	Item	Loading factor	Error Varians	Construct Reliability
KD	KD1	1,00	0,22	0,37
	KD2	0,77	0,38	0,17
	KD3	0,93	0,39	0,23
	KD4	0,86	0,35	0,22
	KD5	0,94	0,29	0,29
Indicator	Item	Loading factor	Error Varians	Construct Reliability
WT	KD6	1,05	0,54	0,21
	WT1	1,00	0,63	0,09
	WT2	1,37	0,65	0,15
	WT3	0,93	0,61	0,08
PD	WT 4	1,33	0,48	0,18
	PD1	1,00	0,62	0,07
PO	PD2	1,30	0,52	0,14
	PO1	1,00	0,49	0,31
	PO2	1,33	0,40	0,50
FT	PO3	0,65	0,87	0,10
	FT1	1,00	0,40	0,45
	FT2	0,24	0,70	0,03
PT	FT3	0,42	0,63	0,08
	FT4	0,46	0,48	0,13
	FT5	0,73	0,53	0,25
	PT1	1,00	0,82	0,21
	PT2	0,95	0,55	0,27
KT	PT3	1,03	0,40	0,37
	PT4	0,81	0,68	0,18
	KT1	1,00	0,46	0,41
	KT2	0,72	0,83	0,17
	KT3	-0,22	0,95	0,02
	KT4	0,60	0,73	0,14

Based on the researchers' opinion that if the loading factor score is $> 0,7$ then the item is fulfill the good criteria (Ghozali & Fuad, 2014). From the loading factor score or convergent validity in the table 4, there are 22 good statement items. Those items are explained as: self-ability indicator (*indikator kemampuan diri – KD*), all items is valid; the time in finishing the assignment indicator (*indikator waktu penyelesaian tugas – W*), all items is valid; self-assessment (*penilaian diri – PD*), all statement items is valid; others-assessment indicator (*indikator penilaian aorang lain – PO*), two items are valid; the focus in finishing the assignment indicator (*indikator fokus dalam menyelesaikan tugas – FT*), two statement items are valid; the mastery of the

assignment (*penguasaan tugas – PT*), all items is valid; and the certain condition indicator (*indikator kondisi tertentu – KT*), has two valid statement items.

Overall, the highest validity in the level dimension is the time in finishing the assignment (WT). The statement item WT2, which is: I do the mathematics question more than one hour (*saya mengerjakan soal matematika lebih dari satu jam*). For the strength dimension, the highest validity is others-assessment indicator (PO), which is PO2, the statement item is my friend says that I am clever in mathematics (*teman saya mengatakan bahwa saya pandai matematika*). While the generality dimension, the highest validity is the mastery of the assignment indicator (PT), which is PT3. The item is I like doing every kind of mathematics question (*saya senang mengerjakan bermacam-macam soal matematika*).



Picture 1. The Result of Confirmatory Factor Analysis

The most reliable item for each indicator can be seen from the construct reliability. The highest construct reliability score is the most reliable item in the indicator. As for those each items in the: KD1 = 0,37; WT4 = 0,18; PD2 = 0,14; PO2 = 0,50; FT1 = 0,45; PT3 = 0,37 and KT1 = 0,41. The summary of the confirmatory factor analysis result, is obtained as in the table 5.

Based on the *fit* indicators of a research model in the table 5, then it is obtained the chi-square score at 689,73 with the probability 0,00, GFI = 0,80 and AGFI = 0,76. The three of them are less fulfill the criteria. But, the RMSEA result analysis shows score 0,074.

Table 5. The Summary of the Confirmatory Factor Analysis Result

<i>Fit</i> Indicator	<i>Fit</i> Model Criteria	Result
Chi-Square dan Probabilitas	$p \geq 0,05$	689,73 ($p = 0.0$)
RMSEA	$< 0,08$	0,074
GFI	0 (tidak <i>fit</i>) – 1 (<i>fit</i>)	0,80
AGFI	0 (tidak <i>fit</i>) – 1 (<i>fit</i>)	0,76

Based on the *fit* indicators of the research model on the table 5, then it is obtained the chi-square score 689,73 with the probability 0,00, GFI = 0,80 and AGFI = 0,76. The three of them is less fulfill the criteria. But, the RMSEA result analysis shows score 0,074. According to Ghazali and Fuad (2014), the thing that is needed to be noticed in seeing the *fit* indicator is RMSEA score because it is more informative. Next, RMSEA acceptance rate based on the criteria is RMSEA score $< 0,080$ (Risnawita, 2013). The interpretation from the table is if it seen from the RMSEA score = 0,074 then it can be concluded that the model is *fit* enough and has a reasonable error. It means that the data obtained from measuring using mathematic self-efficacy scale is appropriate enough with the theory. In other words, the research model has a good accuracy in assessing *fit* model. The result of the research is appropriate with

Risnawati's (2013) research that the result of confirmatory factor analysis of mathematic self-efficacy scale with the *fit* criteria model RMSEA = 0,070 that means the mathematic self-efficacy scale has been developed appropriate with the theory.

CONCLUSION

The requirement that should be fulfilled to make the mathematic self-efficacy scale instrument, inter alia are the content and the construct validity condition. From the result of the empiric validation, it is concluded that: (1) the result of the content validity analysis is using V Aiken coefficient from three people raters with the 40 statement items are 28 items are invalid. It means there is a suitability between the statement items with the mathematic self-efficacy indicator. (2) The result of the construct validity using the LISREL program is avowed that 22 from 28 statement items are good. It means the statement items is appropriate with the mathematic self-efficacy indicator.

The result of empirical test of the content validity and the construct validity of mathematic self-efficacy scale can be a reference to other researchers in making a development of instrument non-test research. However, there is a limitation in this research, which is a comparison of the number of the appropriate item with the mathematic self-efficacy indicator is still less proportional. Especially on the indicator of self-assessment (PD) that only represented by two statement items. So that, it might affect to the research result. Therefore, for the next research, it needs to be done on: (1) the more statement items, (2) the more respondents, (3) other technique of construct validity analysis, such as through *Rasch* model.

REFERENCES

- Aiken, L. R. (1985). Three Coefficients For Analyzing The Reliability And Validity Of Ratings. *Educational And Psychological Measurement*, 45, 131–142.

- <https://doi.org/https://doi.org/10.1177/2F0013164485451012>
- Alam, S. (2018). Apa Itu Mathematics Self-Efficacy? In *Prosiding Seminar Nasional* (Vol. 04, pp. 269–277). Retrieved from <https://www.journal.uncp.ac.id/index.php/proceeding/article/view/1260>
- Alwi, I. (2012). Kriteria Empirik Dalam Menentukan Ukuran Sampel. *Jurnal Formatif*, 2(2), 140–148. <https://doi.org/http://dx.doi.org/10.30998/formatif.v2i2.95>
- Andrew, S. (1998). Self-efficacy as a predictor of academic performance in science. *Journal Of Advanced Nursing*, 27(3), 596–603. <https://doi.org/https://doi.org/10.1046/j.1365-2648.1998.00550.x>
- Arifin, S., & Asdar, A. R. (2015). Profil Pemecahan Masalah Matematika Siswa Ditinjau dari Gaya Kognitif dan Efikasi Diri Pada Siswa Kelas VIII Unggulan SMPN 1 Watampone. *Daya Matematis: Jurnal Inovasi Pendidikan Matematika*, 3(1), 20–29. <https://doi.org/https://doi.org/10.26858/jds.v3i1.1313>
- Azwar, S. (2018). *Penyusunan Skala Psikologi*. Edisi 2. Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2018a). *Reliabilitas dan Validitas*. Edisi 4. Yogyakarta: Pustaka Pelajar.
- Bandura, A. (1986). The Explanatory And Predictive Scope Of Self-Efficacy Theory. *Journal of Social and Clinical Psychology*., 4(i), 359–373. <https://doi.org/https://doi.org/10.1521/jscp.1986.4.3.359>
- Bandura, A. (1997). *Self-Efficacy, The Exercise of Control*. New York: W.H. Freeman and Company.
- Bashoor, K., & Supahar. (2018). Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis Stem. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/http://dx.doi.org/10.21831/pep.v22i2.19590>
- Chao, C. N. G., McInerney, D. M., & Bai, B. (2019). Self-efficacy and Self-concept as Predictors of Language Learning Achievements in an Asian Bilingual Context. *The Asia-Pacific Education Researcher*, 28(2), 139–147. <https://doi.org/10.1007/s40299-018-0420-3>
- Febriana, R. (2015). Uji Validitas Konstruk Pada Instrumen PASS (Procrastination Assessment Scale For Student) Dengan Metode Confirmatory Factor Analysis (CFA). *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, IV(3), 267–277. Retrieved from <http://psikologi.uinjkt.ac.id/wp-content/uploads/2016/08/2-JP3I-Vol.-IV-No.-3-Juli-2015.pdf>
- Firdaningsih. (2016). Efikasi Diri dan Motivasi Berprestasi Dalam Meningkatkan Hasil Belajar Sejarah. *JPPP | Jurnal Penelitian Dan Penilaian Pendidikan*, 1(1), 47–60. <https://doi.org/10.22236/JPPP>
- Ghozali, I & Fuad. (2014). *Structural Equation Modeling, Teori, Konsep dan Aplikasi dengan Program LISREL 9.10*, Edisi 4. Semarang: UNDIP.
- Ghufron, M. N., & Suminta, R. R. (2013). Efikasi Diri dan Hasil Belajar Matematika: Meta-Analisis, 21(1), 20–30. <https://doi.org/https://doi.org/10.22146/bpsi.9843>
- Guyen, B., & Cabakcor, B. O. (2013). Factors in fluencing 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>
- Hartati, L. (2013). Pengaruh Gaya Belajar Dan Sikap Siswa Pada Pelajaran Matematika Terhadap Hasil Belajar Matematika. *Jurnal Formatif*, 3(3), 224–235. <https://doi.org/http://dx.doi.org/10.30998/formatif.v3i3.128>
- Istiqamah, I. (2019). Validitas Perangkat

- Pembelajaran Konsep Pencemaran Lingkungan untuk Meningkatkan Keterampilan Komunikasi Sains Siswa SMA di Banjarmasin. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 8(2), 117. <https://doi.org/10.18592/tarbiyah.v8i2.3235>
- Kusrini., Manoy, J. T., Susanah., & Wijayanti, P. (2014). *Strategi Pembelajaran Matematika*. Edisi 2. Tangerang Selatan: Universitas Terbuka.
- Mardapi, D. (2012). Pengukuran penilaian dan evaluasi pendidikan. *Yogyakarta: Nuha Medika*, 45.
- Naga, D. S. (2013). *Teori Skor Pada Pengukuran Mental*. Jakarta: Nagrani Citrayasa.
- Purnama, A., Wijaya, T. T., Dewi, S. N., & Zulfah, Z. (2020). Analisis Buku Siswa Matematika SMA dari Indonesia dan China Pada Materi Peluang dan Statistik. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 4(2), 813–822. <https://doi.org/https://doi.org/10.31004/cendekia.v4i2.305>
- Risnawita, R. S. (2013). Validitas Konstruk Efikasi Diri Matematika: Analisis Faktor Konfirmatori. *Quality, Journal of Empirical Research In Islamic Education*, 1(2), 71–85. <https://doi.org/http://dx.doi.org/10.21043/quality.v1i2.214>
- Rosyida, F., Utaya, S., & Budijanto. (2016). Pengaruh Kebiasaan Belajar dan Self-Efficacy Terhadap Hasil Belajar Geografi di SMA. *Jurnal Pendidikan Geografi*., 21(2), 17–28. Retrieved from <http://journal.um.ac.id/index.php/pendidikan-geografi/index>
- Santrock, J. W. (2007). Remaja edisi 11 jilid 1. *Jakarta: Erlangga*.
- Sunaryo, Y. (2017). Pengukuran Self-Efficacy Siswa dalam Pembelajaran Matematika di MTs N 2 Ciamis. *Teorema: Teori Dan Riset Matematika*, 1(2), 39–44. <https://doi.org/http://dx.doi.org/10.25157/teorema.v1i2.548>
- Widodo, P. B. (2006). Reliabilitas dan Validitas Konstruk Skala Konsep Diri Untuk Mahasiswa Indonesia. *Jurnal Psikologi Universitas Diponegoro*, 3(1), 1–9.
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23. <https://doi.org/10.18592/tarbiyah.v7i1.2100>