

Development of math efficacy scale for junior high school student in Indonesia

Kemas Rahmat Mubarrak *, Helli Ihsan, Diah Zaleha Wyandini

Department of Psychology, Universitas Pendidikan Indonesia, West Java, Indonesia

* Correspondence: kemasrah@upi.edu

© The Author(s) 2022

Abstract

Indonesian students are indicated to have a low math efficacy. Currently, no psychology scale has been created to measure the math efficacy of Indonesian students, especially for junior high school level students. This study aims to develop a math efficacy scale for junior high school level students or equivalent based on Indonesian students' characteristics. The study used an exploratory sequential design with mixed methods. The results of qualitative analysis through focused group interviews on two small groups of eight students show that seven major themes were related to mathematics efficacy. Qualitative analysis as the basis for a scale development consisting of 33 items and is administered to 478 participants. The results of the quantitative analysis through validity test with Principal Component Analysis (PCA) shows that four dimensions appeared with total variance explained reaching 60.4%. The model was re-tested for compatibility with Confirmatory Factor Analysis (CFA) and obtained index values of $p < 0.001$ (χ^2), 0.047 (SRMR), 0.907 (TLI), 0.918 (CFI), and 0.064 (RMSEA). The four dimensions have met the standard of a fit index with 23 items remaining. The validity test was also supported by Pearson's Product Moment correlation of 0.795 for convergent validity test with Math Attitude Scale (MAS) and divergent validity test of -0.331 with Mathematical Self-Efficacy and Anxiety Questionnaire (MSEAQ). The scale's reliability was very good, with Cronbach's alpha value of 0.918.

Keywords: factor analysis; math efficacy; mixed-method; psychometrics; scale development

Received: 20 November 2021 | Revised: 9 December 2021

Accepted: 29 December 2021 | Published: 6 January 2022



Introduction

Mathematics is a fundamental concept that humans learn from an early age and is essential in their development period (Gopnik et al., 2001; Harris & Petersen, 2017). One important aspect of mathematical concepts is its relation to self-efficacy. Albert Bandura was the first psychologist to introduce self-efficacy in clinical, social, and counseling settings (Zakariya et al., 2019). Bandura (1997) defines perceived self-efficacy as "belief in one's capabilities to organize and execute the courses of action required to produce given attainments." In students, self-efficacy can encourage individuals to develop their abilities and achieve optimal academic achievement (Somawati, 2018). Although Bandura (2012) explains that self-efficacy is not limited to specific tasks, various studies put math efficacy as a single domain. It is also known as math efficacy. Math efficacy can be defined as a person's assessment of their ability to solve a particular mathematical problem (Bonne & Lawes, 2016).

The construct of math efficacy is an aspect that needs to be studied further, as the phenomena that occur indicate that the math efficacy of Indonesian students is relatively low. It is supported by Thien et al.'s (2015) study, which showed that math efficacy is the strongest predictor affecting the performance of Indonesian students in PISA. Indonesia's PISA results for mathematics subject in 2018 put Indonesia in notably low degree in mathematics, which is ranked 73rd out of 79 collaborating countries with an average score of 379. (Schleicher, 2019). Despite the lack of performance of Indonesian students in math subject, research related to math efficacy and its measurement is needed to find out the profile of mathematics efficacy of students in Indonesia.

The Mathematics Self-Efficacy Scale (MSES) developed by Betz and Hackett (1983) was the most widely used scale in measuring math efficacy. Although MSES shows good reliability and stability, its factor structure has never been examined. Later, Kranzler and Pajares (1997) developed the Mathematics Self-Efficacy Scale-Revised (MSES-R) by analyzing factors from MSES. In addition, there are still various scales related to math efficacy. However, most of the other scales were made for college students. Pampaka and Williams (2010) have users limited only to students who are entering a transition to college. Similarly, the scale by Zakariya et al. (2019) only measures mathematical efficacy variables in one particular field of mathematics, namely calculus.

One of the math efficacy scales developed in Indonesia by Sukoco et al. (2018) is Mathematics Self-Efficacy for Senior High School or (MSESc). The scale refers to the previous research of Betz & Hackett (1983). MSESc is made based on the 2015 National Examination grid for the senior high school level. As development research, MSESc has produced a valid and reliable measuring instrument. However, there are still some limitations to this research. The number of samples is still limited and relatively small, with a sample size of only 65 high school students from science and social studies majors. Likewise, the results of factor analysis showed differences in the number of factors in the sample of students majoring in Science and Social Sciences. A scale requires at least 200 data samples to reduce the error rate, and statistically, the measuring instrument should reduce the number of factors with balanced weighting (Coaley, 2010).

This study tried to answer the problems that have been described above by developing a math efficacy scale that follows the education system and the characteristics of Indonesian students. It was done to increase the accuracy of the measuring scale by adjusting the sample characteristics through mixed-method research. With a valid and reliable scale, the assessment should be more effective to help a teacher and every stakeholder get a comprehensive picture of Indonesia's student math efficacy profile.

Methods

The scale was developed based on Bandura's (1997) self-efficacy theory and referred to Mathematics Self-Efficacy Scale-Revised (MSES-R) scale by Kranzler and Pajares (1997). MSES-R has three dimensions, namely solution of math problems (problems), completion of math tasks used in everyday life (tasks), and satisfactory performance in college courses that require knowledge of mathematics (courses) and consists of 52 items. The dimensions used in the MSES-R were re-explored with a qualitative method through a focused group interviews module consisting of 15 questions.

This research used a mixed-method approach with a sequential exploratory research design. Mixed-method aimed for obtaining more comprehensive data by two methods, a qualitative method by a focus group interview and a quantitative method by psychometric test. The development of measuring instruments with mixed-method exploratory sequential design is done through stages. Zhou (2019) divides the process into the following stages: (1) qualitative exploration of the scale construct, which is also a qualitative validation process to collect evidence of content validity; (2) converting qualitative data results into scale items; (3) perform mixing validation to review the validity of item; (4) create scale items and determine item responses; and (5) perform quantitative validation and reliability estimation to analyze the psychometric aspects of the scale. If there was an item with poor psychometric quality, the item must be revised and returned to step (3) for validation tests and re-testing for analysis in step (5).

Participants selected in this study were active students of junior high school level or equivalent in grades 1, 2, and 3 from all over Indonesia. The total number of participants in the study was 486 students, who were divided into 8 participants (4 males and 4 females) for qualitative data and 478 participants (199 male and 279 female) for quantitative data. In qualitative data, the sampling technique used is purposeful sampling with concept sampling methodology. Quantitative data used non-probability sampling techniques with convenience sampling methodology. The number of quantitative samples in this study will be based on factor analysis methods by considering the communality level of the variable-to-factor ratio. Because the data has a wide commonality with a variable-to-factor ratio of 6, the minimum sample required is 200 (Mundfrom et al., 2005).

Qualitative participants consisted of 8 students who were grouped into two groups based on academic grades in school. Group 1 consists of students with high academic scores, and group 2 consists of students with average scores. Participants are in the second grade of junior high school, with ages ranging from 14-15 years. The selection of qualitative participants was

determined directly by the mathematics teacher at one of the public schools in Bandung city. Quantitative participants in this study came from various regions in Indonesia, including Bali, Central Java, Kalimantan, Jakarta, surrounding areas, and West Java, with the dominant number specifically from Bandung city. The participants consisted of male students and female students in grades 1, 2, and 3 of junior high school. The age range of participants for junior high school students is 11-16 years old.

The psychometric quality of the scale was tested using factorial validity with Principal Component Analysis (PCA) and Confirmatory Factor Analysis (CFA) methods to obtain dimensions based on theoretical constructs that adapt to the characteristics of junior high school students in Indonesia. Convergent validity was obtained through comparison with the Math Attitude Scale (MAS) scale developed by Facultad and Sebial (2019). The MAS scale was used as a comparison scale because it has a significant positive correlation (Kundu & Ghose, 2016). Divergent validity was tested with the Mathematical Self-Efficacy and Anxiety Questionnaire (MSEAQ) developed by May (2009). MSEAQ was used because math efficacy was negatively correlated with mathematics anxiety (Akin & Kurbanoglu, 2011). Internal consistency reliability estimation used Cronbach's alpha method. Qualitative data was processed using MAXQDA 2020 software, while quantitative data was processed using IBM SPSS Statistics 26 and JASP 0.14.1.0 software.

In figure 1, below is a chart that contains the stages of the scale development procedure modified from Zhou (2019).

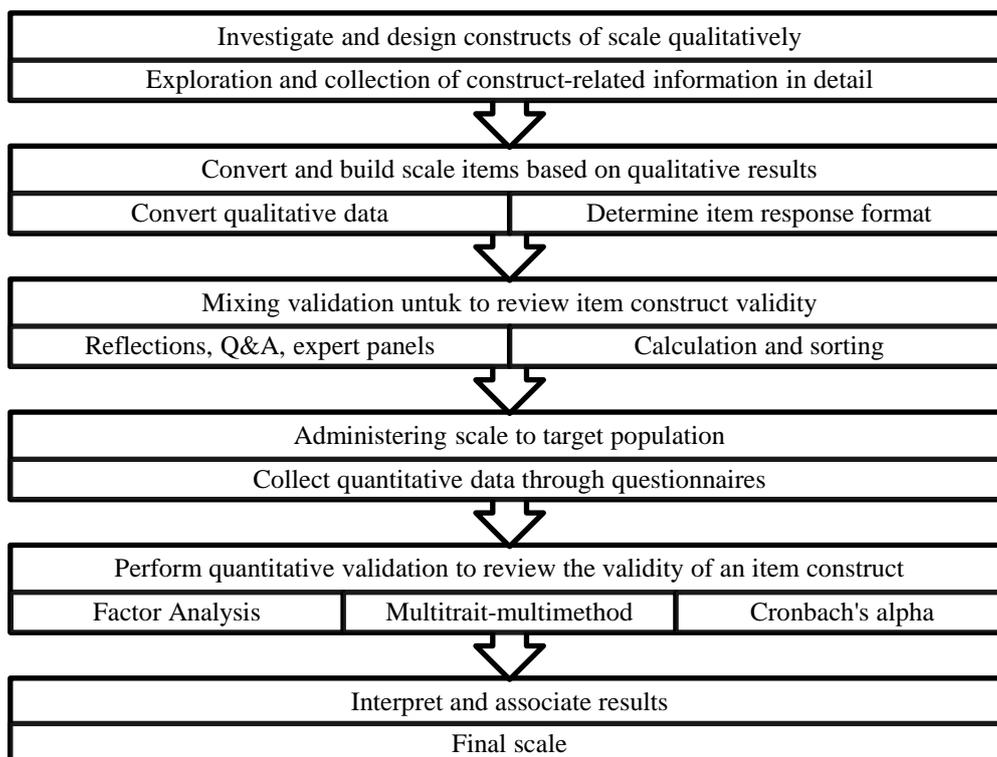


Figure 1. Stages of scale development procedures

Results

Qualitative Analysis

The results of the analysis of the interviews showed that seven themes emerged and could be further understood through the code and its subcodes. The seven themes include feelings that arise, a math course, interpersonal relationships, subjects related to mathematics, maths application, math comprehension, and how to learn mathematics. Based on the analysis results, the frequency of the codes varies and does not always appear in each group. There were no specific differences between the two groups. Although group 1 has academically superior grades, the participants' learning experience does not necessarily indicate a prominent self-confidence in participants both in school and daily activities.

Based on the results of the qualitative analysis, each theme that appears is formulated as a dimension first. The items in each dimension are constructed based on the code and the citations obtained. At this stage, the items from the initial scale reach more than 50 items. Next, mixing validation was carried out with reflection, question and answer, and expert panel review. Reflection is done by making an operational definition of the construct measured on the scale. The main construct that is measured is the math efficacy of an individual's level of confidence to complete mathematical activities in the context of learning at school and its application in everyday life. At the question-and-answer stage and expert panel review, item screening was carried out by expert panel judgment based on the item's ability to represent constructs. The final result in the qualitative stage was an initial scale with 33 items without dimensions. These scales then are explored and reduced through quantitative analysis formulated upon the psychometric test of Principal Component Analysis and Confirmatory Factor Analysis.

Item selection based on discriminatory power

The first step in selecting items on the scale was to look at the discriminatory power of the items. Discriminatory power can indicate whether the items have been well structured, meaningful, and have the functions needed to assess the subject's experience (Boateng et al., 2018). Good items will have discriminatory power that can distinguish the subject's characteristics based on the construct measured by the measuring instrument (Azwar, 2019). In other words, the item also has consistency according to its function and the function of the scale. The discriminatory power of items can be seen through the corrected item-total correlations score. Items with a score of 0.30 are considered usable (Azwar, 2019). Based on the analysis results on all items in the scale, five items have poor quality and will be discarded, namely item numbers 2, 4, 17, 18, and 25. The items that were discarded included "I repeatedly counted to answer math questions" and "I depend on various online media (Youtube, Google) to solve math problems." The five items will not be used as part of the scale in the next stage.

Principal Component Analysis

To assess the scale's construct validity, the first step is Principal Component Analysis (PCA). PCA is a mathematical procedure for changing correlated variables into a more straightforward set of variables (Dharmawardena et al., 2017). PCA is a technique for performing reductions in improving the interpretation of a scale model by minimizing missing information (Hair et al., 2014; Jolliffe & Cadima, 2016). In the Table 1 below is the results of the PCA.

Table 1. Result of Principal Component Analysis (PCA)

	Component Loadings				Uniqueness
	Positive View	Negative Affect	Math Application	Out-of-Class Learning	
Item 1	0.784				0.348
Item 3	0.802				0.480
Item 5		0.777			0.369
Item 6	0.622				0.442
Item 7	0.610				0.395
Item 8		0.866			0.259
Item 9		0.836			0.294
Item 11	0.686				0.488
Item 12	0.731				0.405
Item 13	0.709				0.449
Item 14	0.550				0.501
Item 15	0.574				0.549
Item 19			0.624		0.560
Item 20			0.800		0.369
Item 21			0.829		0.293
Item 22			0.791		0.267
Item 26		0.720			0.379
Item 27	0.778				0.397
Item 28	0.602				0.398
Item 29				0.852	0.248
Item 30				0.780	0.271
Item 31	0.570				0.421
Item 32	0.402				0.535

Note. The rotation method applied is oblimin.

The results of the PCA show that there were four factors that arose. Item analysis is carried out at this stage. Items with component loadings < 0.4 will be discarded. From the 28 items, only 23 items remained. A total of 5 items were removed from the scale since they did not meet the component loading value criteria and/or the items happened to be cross-loadings. Examples of discarded items include “I can complete math assignments independently without the help of my brother or sister at home” and “I can calculate the money I have to pay when I am splitting bills with friends.” Factor 1 consisted of 13 items, factor 2 consisted of 4 items, factor 3 consisted of 4 items, and factor 4 consisted of 2 items. These items have component

loadings that are considered high, with a value above 0.5. It can indicate that the item has a good structure and can linearly measure each factor component (Arguello & Crescenzi, 2019; Hair et al., 2014).

Another criterion applied to this method is the percentage of variance explained. The following Table 2 is the result of the accumulated variance explained on 23 items of the scale:

Table 2. Accumulated variance explained

	Eigenvalue	Variance Proportion	Accumulation
Positive View	8.785	0.382	0.382
Negative Affect	2.569	0.112	0.494
Math Application	1.497	0.065	0.559
Out-of-Class Learning	1.031	0.045	0.604

The accumulation of variance explained reaches 60.4%. The variance value explained above 60% is considered satisfactory (Hair et al., 2014). Based on the overall PCA results, the scale has met the criteria well according to the component loadings values, and the variance explained. Furthermore, each factor needs to be tested further through the confirmatory factor analysis method.

Confirmatory Factor Analysis

The next step to assess the scale's construct validity was Confirmatory Factor Analysis (CFA). In CFA, an evaluation of each factor will be carried out to assess the accuracy of the item by estimating the relationship between latent constructs (Boateng et al., 2018; Brown, 2015). The method used in estimating CFA is Maximum Likelihood. Estimates are selected based on a continuum and multivariate data types (Li, 2016). CFA calculations are processed through the JASP 0.14.1.0 program. Figure 2 is a model of the scale based on the previous PCA stages.

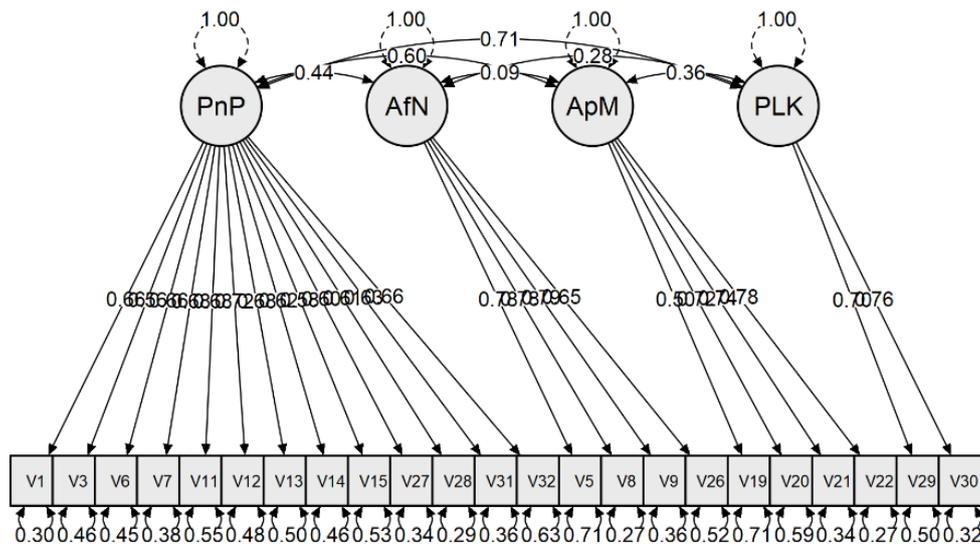


Figure 2. Result of Confirmatory Factor Analysis (CFA)

It can be seen in Figure 2 that there was a relationship between the four factors of the math efficacy scale. To be able to interpret the results of the analysis, CFA used parameters of

fit indices consisted of absolute fit measured, namely Chi-square (χ^2) and Standardized Root Mean Square Residual (SRMR); incremental fit indices namely Tucker Lewis Index (TLI) and Comparative Fit Index (CFI); and a discrepancy index of Root Mean Square Error of Approximation (RMSEA). The standard parameter values used are based on Hair et al., (2014) for the number of samples $N > 250$ and the number of variables is 30, considering that the initial scale has 33 items. In Table 3 below is the results of fit indices of the CFA.

Table 3. Fit indices result of Confirmatory Factor Analysis

Fit Indices	Indices Value	Standard Value
χ^2	$p < .001$	$p < .001$
SRMR	0.047	≤ 0.08
TLI	0.907	> 0.90
CFI	0.918	> 0.90
RMSEA	0.064	< 0.07

From the table 3, it can be seen that the statistical test of the CFA model on 23 scale items in 4 factors is considered to have met the standard value for each fit indices' parameters. For absolute fit index, p-value has significance with $p < .001$. The SRMS value obtained is 0.045 which is lower than 0.08 and is in accordance with the standard. The two incremental indices, both TLI and CFI, have values that are close to the standard, namely > 0.90 . Therefore, the two incremental indices are considered to be fit, although not satisfactory. In the discrepancy indices, the RMSEA value with value of 0.064 has met the standard < 0.07 . Overall, the factor model of the scale is in accordance with the standard of fit indices. In Table 4, below is the final items and dimension of scale:

Table 4. Final items and dimension scale of AUKEMI

Dimension	Indicator	Item
Positive View	Individuals have positive and constructive feelings, beliefs, and activities in learning mathematics.	1. <i>Saya yakin memiliki kemampuan yang baik dalam pelajaran matematika</i>
		2. <i>Saya yakin terhadap jawaban saya ketika menghitung soal matematika</i>
		3. <i>Saya merasa senang ketika mengerjakan soal-soal matematika</i>
		4. <i>Saya merasa senang untuk mempelajari matematika</i>
		5. <i>Saya merasa puas terhadap kemampuan saya dalam bidang matematika</i>
		6. <i>Saya merasa mampu untuk bersaing dengan siswa lain pada mata pelajaran matematika</i>
		7. <i>Saya percaya diri dapat memperoleh nilai matematika yang lebih tinggi dibandingkan teman saya yang lain</i>
		8. <i>Saya mampu menjawab soal matematika secara mandiri tanpa bantuan teman</i>
		9. <i>Saya mampu memberikan bantuan pada teman yang kesulitan belajar matematika</i>
		10. <i>Saya memperoleh nilai yang baik di berbagai tugas dan ulangan matematika</i>

Dimension	Indicator	Item
		11. <i>Saya mampu memahami berbagai materi matematika</i>
		12. <i>Saya dapat menerapkan dengan cepat materi yang baru diajarkan oleh guru matematika</i>
		13. <i>Saya merasa percaya diri dengan bertanya pada guru saat mempelajari matematika yang sulit</i>
Negative Affect	Individuals are not or minimally influenced by negative emotions in learning mathematics.	14. <i>Saya merasa matematika adalah pelajaran yang menakutkan</i>
		15. <i>Saya merasa malas untuk mempelajari matematika</i>
		16. <i>Saya merasa malas untuk menghafalkan berbagai rumus dalam mata pelajaran matematika</i>
		17. <i>Saya merasa soal-soal matematika itu susah untuk dikerjakan</i>
Math Application	Individuals feel capable of carrying out various activities that require the application of the field of mathematics in everyday life.	18. <i>Saya menggunakan media online sebagai referensi tambahan untuk belajar matematika</i>
		19. <i>Saya mampu untuk berbelanja kebutuhan pribadi di toko swalayan</i>
		20. <i>Saya mampu untuk menghitung dan mengatur uang yang saya gunakan sehari-hari</i>
		21. <i>Saya mampu untuk menghitung uang yang harus saya bayarkan saat berpatungan dengan teman</i>
Out-of-Class Learning	Individuals intentionally carry out various activities outside of formal school hours to develop their mathematical abilities	23. <i>Saya memiliki waktu belajar tambahan di luar kelas untuk mempelajari matematika</i>
		24. <i>Saya mengerjakan latihan soal matematika di luar tugas maupun ulangan di sekolah</i>

Convergent and divergent validity

Convergent and divergent validity is a construct validity test in the multitrait-multimethod approach. In the convergent validity test, the total score of the math efficacy scale with 23 items was compared with Math Attitude Scale (MAS) scale. Meanwhile, in the divergent validity test, the math efficacy scale was compared with the Mathematical Self-Efficacy and Anxiety Questionnaire (MSEAQ). In the MSEAQ scale, only anxiety items are used as a comparison.

Prior to the validity test at this stage, the two comparative scales had been analyzed for reliability and were classified as having satisfactory reliability. The MAS scale with Cronbach's alpha value of 0.904 is considered as having good reliability. While in the MSEAQ Cronbach's alpha value obtained is 0.954 and is also considered very good. Furthermore, the math efficacy scale was carried out with a correlation test using the Pearson method. In Table 5, below is the results of the correlation test.

Table 5. Correlation test results of math efficacy scale

	<i>Pearson's</i>	
	r	P
Math Efficacy - MAS	0.795	< .001
Math Efficacy - MSEAQ	-0.331	< .001

Based on table 5 above, it can be seen that the convergent validity correlation test has a strong correlation with $r > 0.5$. The r-value obtained is 0.795, indicating that the MAS scale and math efficacy have a positive relationship. The p-value < 0.001 indicates that the correlation between the two scales is significant. It can be understood that the constructs of efficacy and mathematics attitudes have a positive and related relationship. The divergent validity test between the MSEAQ scale and math efficacy shows a negative relationship between the two, namely -0.331 with a significance of $< .001$. It shows that the two scales even measure two different constructs but have a relationship between efficacy and math anxiety.

Factor naming

At this stage, the math scale was valid and reliable based on psychometric tests. The naming of factors is done after the CFA process is complete based on the results of the analysis and grouping the meanings of the items formed by the CFA. To make it easier to pronounce, the scale will be named Alat Ukur Efikasi Matematika Indonesia (AUKEMI). In this scale there are four factors or dimensions, namely positive view, negative affect, math application, and out-of-class learning. The following is an operational definition of each factor math efficacy scale:

- a. Positive View: an individual's self-view of positive and constructive feelings, beliefs, and activities in relation to all aspects of the field of mathematics.
- b. Negative Affect: all negative feelings and emotions felt by individuals in relation to all aspects of mathematics.
- c. Math Application: an individual's self-assessment on his ability to perform various activities that require the application of the field of mathematics in everyday life.
- d. Out-of-Class Learning: various activities that are intentionally carried out by individuals to develop their mathematical abilities outside of formal school hours.

Reliability

Table 6. Reliability of Alat Ukur Efikasi Matematika Indonesia (AUKEMI)

Dimension	Number of Item	Alpha
Positive View	13	0.924
Negative Affect	4	0.829
Math Application	4	0.791
Out-of-Class Learning	2	0.718
Math Efficacy (AUKEMI)	23	0.918

Based on the results of the reliability estimation in Table 6 above, it is found that the overall reliability of the math efficacy scale with 23 items that have been validated is very good, with a value of 0.918. However, from each scale dimension, the reliability value includes wide interpretations with moderate to very good interpretations. Dimensions of Math Applications and Out-of-Class Learning with an alpha value in the range of $0.71 < x < 0.80$ considered has moderate reliability. The Negative Affect dimension is considered good, and the Positive View dimension is considered very good. It should be understood that the difference in the number of items in each dimension will affect the reliability. The more items, the higher the reliability produced (Kaplan & Saccuzzo, 2013).

Discussion

The psychometric aspect of the scale follows the standard. Compared with the Mathematics Self-Efficacy Scale-Revised (MSES-R) (Kranzler & Pajares, 1997), which is used as a reference, several differences are developed in this scale. The dimensions of the MSES-R, which are math problems, tasks, and courses, have been developed into four dimensions, namely positive view, negative affect, math application, and out-of-class learning. In contrast to the MSES-R, which emphasizes various mathematical activities and tasks, AUKEMI adds aspects in the form of views and feelings experienced by individuals. An item with concrete and specific mathematical questions in MSES-R is not used in AUKEMI. Items in AUKEMI gave an overall view of individuals on their ability to answer math problems without any specific form of the mathematics question to solve. It is in line with Bandura's (2012) view that the strength of self-efficacy should be measured across various performances in a broad activity domain, not just a specific performance domain.

In the psychometric aspect, MSES-R has a more satisfactory reliability value. The MSES-R reliability value with Cronbach's alpha is very good, with a value of 0.95. Likewise, the reliability of each dimension is more stable, with a value of 0.94 for math tasks, 0.91 for math courses, and 0.91 for math problems. In terms of validity, MSES-R, which uses the Principal Component Analysis method, only obtained a total variance explained value of 55.3% compared to AUKEMI of 60.4%. With the explained variance value above 60%, AUKEMI is classified as satisfactory (Hair et al., 2014). However, it should also be understood that MSES-R was tested on a larger sample with 522 participants compared to AUKEMI, which only had 478 participants.

Although AUKEMI has met the psychometric standards of measuring instruments, some aspects can still be developed. First, the measuring instrument, which is now structured, needs to be re-administered on larger and more samples. It can support the results of the CFA test, which so far are still based on the same sample when testing PCA. Another development that can be done is to add items to the dimensions with relatively few items to increase its reliability (Kaplan & Saccuzzo, 2013). For example, the dimension of out-of-class learning with only two items should have more items. It takes at least three items with high factor loading values in each dimension to represent the dimensions well (Raubenheimer, 2004). Dimensions on the scale that have been defined operationally should make adding items easier for the next researcher.

Conclusion

The mathematical efficacy constructs measured in Alat Ukur Efikasi Matematika Indonesia (AUKEMI) consists of 4 dimensions that are considered adequate for defining math efficacy in individuals, namely positive views, negative affect, mathematical applications, and learning outside the classroom. In terms of psychometrics, AUKEMI shows fair factorial validity and meets the model criteria. Similarly, the convergent validity with Mathematics Attitude Scale (MAS) and divergent validity with the Mathematics Self-Efficacy and Anxiety Questionnaire (MSEAQ) show a consistent relationship. In terms of reliability, the scale has a high-value coefficient based on Cronbach's alpha and is considered favorable. Overall, the AUKEMI scale has good validity and reliability and can represent the level of math efficacy of individuals.

Conflicts of Interest

Authors declare there is no conflict of interest regarding the publication of this manuscript.

References

- Akin, A., & Kurbanoglu, I. N. (2011). The relationships between math anxiety, math attitudes, and self-efficacy: A structural equation model. *Studia Psychologica*, 53(3), 263–273. <https://psycnet.apa.org/record/2011-27554-004>
- Arguello, J., & Crescenzi, A. (2019). Using principal component analysis to better understand behavioral measures and their effects. *ICTIR 2019 - Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, 177–184. <https://doi.org/10.1145/3341981.3344222>
- Azwar, S. (2019). *Penyusunan skala psikologi [Making psychological scale]* (2nd ed.). Pustaka Pelajar.
- Bandura, A. (1997). *Self-efficacy the exercise of control*. W. H. Freeman and Co.
- Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, 38(1), 9–44. <https://doi.org/10.1177/0149206311410606>
- Betz, E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior*, 23(3), 329–345. [https://doi.org/10.1016/0001-8791\(83\)90046-5](https://doi.org/10.1016/0001-8791(83)90046-5)
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L.

- (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bonne, L., & Lawes, E. (2016). Assessing students' maths self-efficacy and achievement. *Research Information for Teachers (Set)*, 2, 60–63. <https://doi.org/10.18296/set.0048>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Coaley, K. (2010). *An introduction to psychological assessment and psychometrics*. SAGE Publications.
- Dharmawardena, P., Thattil, R. O., & Samita, S. (2017). Scale independent principal component analysis and factor analysis with preserved inherent variability of the indicators. *American Journal of Theoretical and Applied Statistics*, 6(2), 90-94. <https://doi.org/10.11648/j.ajtas.20170602.13>
- Facultad, E. B., & Sebial, S. C. L. (2019). Development and validation of a mathematics attitude scale (MAS) for high school students in Southern Philippines. *International Journal of Innovation, Creativity and Change*, 8(7), 147–162. <https://doi.org/10.2139/ssrn.3503680>
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5), 620–629. <https://doi.org/10.1037//0012-1649.37.5.620>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. In *Pearson Education Limited* (7th ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_395
- Harris, B., & Petersen, D. (2017). Developing math skills in early childhood. In *Mathematica Policy Research* (Issue August). <https://files.eric.ed.gov/fulltext/ED587415.pdf>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065), 1–16. <https://doi.org/10.1098/rsta.2015.0202>
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological testing: Principles, applications, and issues*. In *Wadsworth* (8th ed.).
- Kranzler, J. H., & Pajares, F. (1997). *An exploratory factor analysis of the mathematics self-efficacy scale—revised (MSES-R)*. 29(4), 215–228. <https://doi.org/10.1080/07481756.1997.12068906>
- Kundu, A., & Ghose, A. (2016). The relationship between attitude and self efficacy in mathematics among higher secondary students. *IOSR Journal Of Humanities And Social Science (IOSR-JHSS)*, 21(4), 25–31.
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- May, D. K. (2009). *Mathematics self-efficacy and anxiety questionnaire*. University of Georgia.
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4
- Pampaka, M., & Williams, J. (2010). Measuring mathematics self efficacy of students at the beginning of their higher education studies. *Proceedings of the British Congress for Mathematics Education, January*, 159–166. <http://www.bsrlm.org.uk/wp-content/uploads/2016/02/BSRLM-IP-30-1-21.pdf>
- Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, 30(4), 59–64. <https://doi.org/10.4102/sajip.v30i4.168>

- Schleicher, A. (2019). *PISA 2018 Insights and Interpretations*. OECD Publishing.
- Somawati, S. (2018). Peran efikasi diri (self efficacy) terhadap kemampuan pemecahan masalah matematika [The role of self-efficacy on math problem solving skills]. *Jurnal Konseling dan Pendidikan*, 6(1), 39-45. <https://doi.org/10.29210/118800>
- Sukoco, H., Wismono, A., & Rusmilah. (2018). Skala mathematics self-efficacy untuk jenjang SMA [Mathematics self-efficacy for senior high school]. *Prosiding Seminar Nasional Etnomatnesia*, 752–760. <https://jurnal.ustjogja.ac.id/index.php/etnomatnesia/article/view/2413/1374>
- Thien, L. M., Darmawan, I. G. N., & Ong, M. Y. (2015). Affective characteristics and mathematics performance in Indonesia, Malaysia, and Thailand: What can PISA 2012 data tell us? *Large-Scale Assessments in Education*, 3(3), 1–16. <https://doi.org/10.1186/s40536-015-0013-z>
- Zakariya, Y. F., Goodchild, S., Bjørkestøl, K., & Nilsen, H. K. (2019). Calculus self-efficacy inventory: Its development and relationship with approaches to learning. *Education Sciences*, 9(3), Article 170. <https://doi.org/10.3390/educsci9030170>
- Zhou, Y. (2019). Measurement: Interdisciplinary research and a mixed methods model of scale development and validation analysis. *Measurement: Interdisciplinary Research and Perspectives*, 17(1), 38–47. <https://doi.org/10.1080/15366367.2018.1479088>