A meta-analysis of the relationship between self-assessment and Mathematics learning achievement

Fransidha Sidhara Hadi[1*]; Janu Arlinwibowo[2]; Gupita Nadindra Fatima[3]

[1]Universitas Negeri Yogyakarta, Indonesia

[2]Badan Riset dan Inovasi Nasional, Indonesia

[3]National Yang Ming Chiao Tung University & International College of Semiconductor Technology (ICST), Taiwan

*Corresponding Author. E-mail: fransidhasidharahadi@gmail.com

ARTICLE INFO

ABSTRACT

This study uses a meta-analysis model to describe the relationship between self-assessment and achievement in learning mathematics. This meta-analysis covers articles published from 2011 to 2022 and is restricted to studies published in English. Indexed articles by Google Scholar were chosen. The data must meet these criteria: quantitative research, containing correlational research on the relationship between self-assessment and mathematics learning achievement (including correlation values and sample size). Through the data screening process with predetermined criteria, 12 research results were obtained, containing 43 studies to be analyzed. This meta-analysis uses a random effect model due to the heterogeneous data distribution. A publication bias test was carried out with the Fail-safe N model to ensure the quality of the data. The result of the analysis showed that the data distribution was heterogeneous, according to the I2 test, so selecting the random effect model was the right decision. Regarding publication bias, an accurate Fail-safe N test shows that the data are free from publication bias. Thus, the analysis uses a suitable model, and the results of the analysis can be trusted. The total effect size indicates a significant positive correlation (r = 0.295) between self-assessment and students' mathematics achievement. Therefore, the higher the self-assessment index, the higher one's mathematics learning achievement.

INTRODUCTION

Assessment is an essential and inseparable learning component (Arlinwibowo et al., 2021). The information obtained through the assessment process is used to make decisions about students, learning content, planning, and educational policies (Nitko & Brookhart, 2014). The Indonesian Government established Curriculum 2013 and the Merdeka Curriculum as the current curriculum, and self-assessment has been introduced in each of these curricula. The study about self-assessment has been conducted by researchers and each has emphasized various perspectives about what self-assessment refers to.

Talking about the definition of self-assessment, Brown and Harris (2013) write that self-assessment is a descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities. On the other hand, Yan (2016) defines self-assessment as a skill to evaluate one's knowledge, skills, and performance or as a step of the practice/learning process. Meanwhile, in the context of the existing curriculum in Indonesia, self-assessment is often referred to as "*penilaian diri*". The Ministry of Education and Culture's Curriculum 2013 Assessment Guide states that assessment focuses on learning outcomes and the learning proc-

ess. Students also begin to be involved in "*penilaian diri*" to practice self-assessment. "*Penilaian diri*" is used as a support for attitude assessment. "*Penilaian diri*" is carried out at least once a year at the end of the semester (Ministry of Education and Culture, 2016, p.22).

Self-assessment is included in one of the ideas from a growth mindset. This idea is applied in the assessment, hoping to develop awareness that the final result is just one of many essential things. However, the process behind the final result, namely the achievement of learning objectives, is more important than just the final result (Ministry of Education, Culture, Research, and Technology, 2021). Thus, teachers are expected to be able to familiarize students with self-assessment as part of this effort. Unfortunately, a gap exists between the government's expectations and what teachers face. Only a very few Indonesian teachers are applying self-assessment for their students (Setiadi, 2016). In line with that fact, the self-assessment application is still limited to a formality related to the pressure and burden on teachers to fulfill their administrative duties, and even many classes rarely implement self-assessment (Brown & Harris, 2013). Relating this case, the researcher saw a link between the infrequent implementation of self-assessment in the classroom and the teacher's lack of confidence in the urgency of the self-assessment itself. The teacher needs to get a strong push or "strong why" that self-assessment is essential in learning, not just an empty appeal from the curriculum. Moreover, self-assessment improves academic performance (Yan et al., 2021) and contributes to higher learning achievement (McMillan, 2013; Brown & Harris, 2014). More specifically, there is a positive effect of self-assessment training on students' mathematical achievement as statistically revealed by Popelka (2015) and also Baiduri (2022).

Learning achievement is still one of the benchmarks for the success of the learning process in the classroom. This achievement is easy to see through the various academic scores obtained by students, and it is this learning achievement that teachers have reported to parents and other relevant stakeholders. Several previous meta-analytic studies showed a close relationship between self-assessment and academic achievement/performance, with the median effect size being between $d = 0.40\text{-}0.45$ (Brown & Harris, 2013; Panadero et al., 2014). Likewise, there have been studies on the relationship between self-assessment and mathematics learning achievement. Therefore, the researchers here are interested in analyzing in more depth how far the link between self-assessment and mathematics learning achievement is through a meta-analysis study. Furthermore, the researchers also have a conjecture that education level and region have an effect on self-assessment and achievement in learning mathematics. The researchers hope this article will add insight into self-assessment in mathematics education in general and can help readers, especially teachers, as a first step which can later develop to be interested in further understanding of self-assessment in their mathematics class.

RESEARCH METHOD

The study focuses on articles published in peer-reviewed journals to maintain the quality of the research methodology of the selected articles (Castro-Alonso et al., 2019). The researchers used Google Scholar as the search engine to find relevant articles with the keywords "correlation", "self-assessment", or equivalent terms such as "self-evaluation", "self-reflection", "self-regulated learning", "metacognitive" (Dent & Koenka, 2016; Youde, 2019), "mathematics achievement".

Google Scholar is connected to various journal portals and journal indexing institutions so that searches can be conducted as widely as possible to find articles representing global conditions to avoid bias (Arlinwibowo et al., 2022). The "cited by" feature in Google Scholar is also used to access related studies (Martin-Martin et al., 2017). The selected articles are in English to maintain transparency in the global community (Krieglstein et al., 2022). In addition, selected articles were published in 2011-2022 to keep information up-to-date. Afterward, articles were selected based on the inclusion criteria (Figure 1).
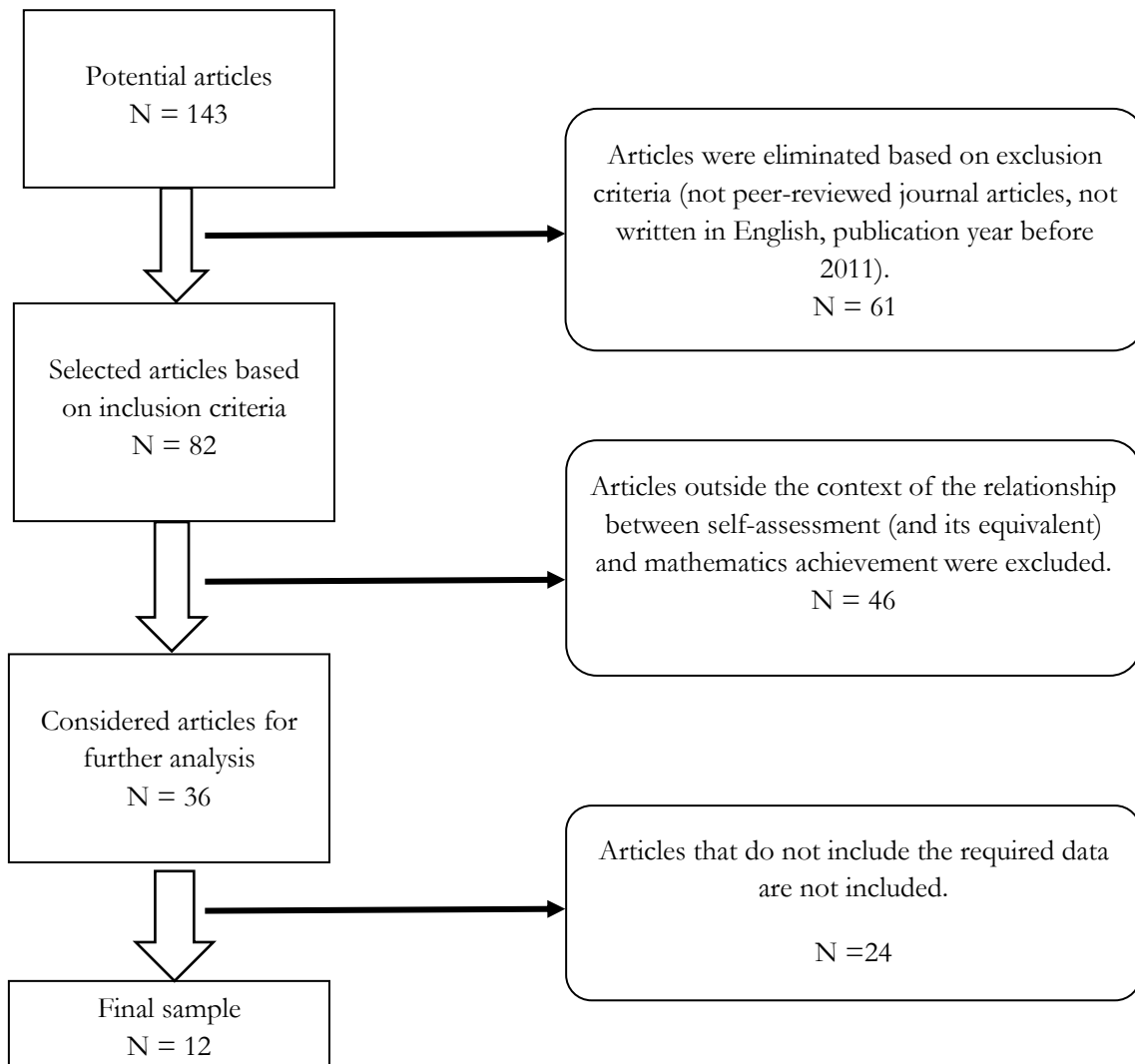
Figure 1. Flowchart of Data Screening

This study uses a random model. The random model was determined based on the research characteristics collected as research data. The data analyzed indicated heterogeneity, as shown by the various levels of education, the research location (country), and the instruments used. We chose the random analysis process based on these characteristics. Furthermore, the value of I^2 clarifies the accuracy of selecting the model. Data are said to be heterogeneous if $I^2 > 25\%$.

We re-examined the selected articles to ensure the self-assessment context is in the mathematics subject area. Therefore, studies that report a correlation (r) will be selected for further analysis as an effect size through meta-analysis (Retnawati et al., 2018). Articles that do not include the correlation coefficient and the number of participants in the study will be excluded from the selected sample unless t-statistical or F data can still be obtained. They will be converted into r using the Formula (1) with $t = \sqrt{F}$:

$$ r = \frac{t}{\sqrt{t^2 + n - 2}} \qquad\dots\dots\dots\dots\dots\dots.. \ (1) $$

where *n* is the number of populations or samples, all included articles will be reviewed to retrieve the necessary data. Correlation is transformed by using the z-Fisher formula, as shown in Formula (2).

$$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right) \ldots\ldots\ldots\ldots\ldots\ (2)$$

Then, to obtain a summary effect, data that were analyzed were converted back into a correlation as in Formula (3).

$$r = \frac{e^{2z}-1}{e^{2z}+1} \ldots\ldots\ldots\ldots\ldots\ (3)$$

These correlations were used to read the final analysis results (Retnawati et al., 2018). The input data are z as the effect size and SEz as the standard error of the effect size. SE$_z$ can be determined in advance using Formula (4).

$$= \sqrt{\frac{1}{(n-3)}} \ldots\ldots\ldots\ldots\ldots\ (4)$$

The data obtained above (z and SEz) were processed to display a forest plot that summarized each study's interval values and standard errors and their conclusions (Arlinwibowo et al., 2022). In addition, to ensure the quality of the data and the conclusions of the analysis results, the indication identification of publication bias was carried out. The conclusions of the meta-analysis results can be trusted if they are free from publication bias. Identification of publication indications can be carried out using the fail-safe N method. This method will produce a value with the criteria if the value of fail-safe N > 5K+10, it can be concluded that no publication bias problem exists. K on these criteria is the number of studies involved in the analysis process.

Moderating variable analysis was carried out by examining the selected studies' descriptor covariance and effect size. In the moderator analysis, the study's effect size becomes the dependent variable, and the research descriptor (in this case, the study level and area) becomes the independent variable (Cooper et al., 2009). Dent and Koenka (2016) state that the variation within each group (study level and region) is tested by meta-analysis using a within-group goodness-of-fit statistic (Qw). The significant Qw statistic indicates a systematic variation among the 43 mean correlations. A between-groups goodness-of-fit statistic (Qb) indicates more variation among the mean correlations for moderator categories than can be explained. Qb value can be obtained using Formula (5):

$$Qb = Q - Qw \ldots\ldots\ldots\ldots\ldots\ (5)$$

Q is the overall correlation, and Qw is the total Q value of each member of the moderator group. The moderator testing analysis related to student education levels was carried out by categorizing students into four groups,-: JHS, SHS, and Higher Education (College) so that Qw = QES + QJHS + QSHS + QCollege. Besides, the region is categorized into East (Eastern) and West (Western). This categorization is related to the finding that many instruments were developed by western countries and turned out to be influential if used by students in the eastern region (Tian et al., 2018; Fadlelmula et al., 2014).

FINDINGS AND DISCUSSION

Findings

A total of 43 studies analyzed through this research came from 12 articles that produced data on the relationship between self-assessment and students' mathematics achievement, as summarized in Table 1. Several studies have produced several different correlations due to the

differentiation of student achievement tests into several types, for example, midterm tests (or many studies call them periodic tests) and end-of-semester tests (Baiduri, 2022; Zimmerman et al., 2011). Some studies differentiate between male and female participants (Altun & Elden, 2013). In addition, several studies include t-values (Cabedo & Maset-Laudes, 2019) and F values (Lai & Hwang, 2016). Studies with an r = 0.5 have a strong relationship, r = 0.3 indicates a moderate relationship, and r = 0.1 indicates a weak one (Li, 2022).

Table 1. Summary Table of Obtained Data

ID	Study	r	N	t	F	Participant	Country	Instrument
	Strong Category (Studies)							
5	Lai & Hwang (2016) Study 2	0.80	11	3.98	15.82	ES	Taiwan	custom
43	Özsoy (2011)	0.648	242			ES	Turkey	MSA-TR
4	Lai & Hwang (2016) Study 1	0.63	11	2.41	5.81	ES	Taiwan	custom
10	Altun & Elden (2013) Study 3	0.59	329			College	Turkey	MSLQ
8	Altun & Elden (2013) Study 1	0.54	472			College	Turkey	MSLQ
	Moderate Category (20 Studies)							
11	Zimmerman et al. (2011) Study 1	0.49	496			College	USA	custom
6	Özcan (2014) Study 1	0.42	408			ES	Turkey	Jr.MAI
15	Zare et al. (2022) Study 1	0.42	278			SHS	Iran	custom
2	Baiduri (2022) Study 2	0.41	114			College	Indonesia	custom
16	Zare et al. (2022) Study 2	0.40	278			SHS	Iran	custom
13	Zimmerman et al. (2011) Study 3	0.39	496			College	USA	custom
12	Zimmerman et al. (2011) Study 2	0.38	496			College	USA	custom
9	Altun & Elden (2013) Study 2	0.36	143			College	Turkey	MSLQ
34	Ahmed et al. (2013) Study 1	0.36	495			JHS VII	Netherlands	MSLQ
1	Baiduri (2022) Study 1	0.35	114			College	Indonesia	custom
35	Ahmed et al. (2013) Study 2	0.35	495			JHS VII	Netherlands	MSLQ
38	Ahmed et al. (2013) Study 5	0.35	495			JHS VII	Netherlands	MSLQ
14	Zimmerman et al. (2011) Study 4	0.34	496			College	USA	custom
17	Rosario et al. (2013) Study 1	0.33	756			JHS	Portugal	custom
41	Ahmed et al. (2013) Study 8	0.33	495			JHS VII	Netherlands	MSLQ
42	Ahmed et al. (2013) Study 9	0.31	495			JHS VII	Netherlands	MSLQ
18	Rosario et al. (2014) Study 2	0.3	756			JHS	Portugal	custom
21	Rosario et al. (2014) Study 5	0.3	756			JHS	Portugal	custom
36	Ahmed et al. (2013) Study 3	0.3	495			JHS VII	Netherlands	MSLQ
39	Ahmed et al. (2013) Study 6	0.3	495			JHS VII	Netherlands	MSLQ
	Low Category (18 Studies)							
40	Ahmed et al. (2013) Study 7	0.29	495			JHS VII	Netherlands	MSLQ
23	Rosario et al. (2014) Study 7	0.27	756			JHS	Portugal	custom
28	Tian et al. (2018) Study 1	0.27	569			SHS	Taiwan	MKMQ
7	Özcan (2014) Study 2	0.26	140			ES	Turkey	Jr.MAI
22	Rosario et al. (2014) Study 6	0.26	756			JHS	Portugal	custom
29	Tian et al. (2018) Study 2	0.21	569			SHS	Taiwan	MKMQ
37	Ahmed et al. (2013) Study 4	0.2	495			JHS VII	Netherlands	MSLQ
19	Rosario et al. (2014) Study 3	0.19	756			JHS	Portugal	custom
30	Tian et al. (2018) Study 3	0.19	569			SHS	Taiwan	MKMQ
20	Rosario et al. (2014) Study 4	0.17	756			JHS	Portugal	custom
24	Fadlelmula et al. (2013) Study 1	0.08	1019			JHS	Turkey	MSLQ
27	Fadlelmula et al. (2013) Study 4	0.06	1019			JHS	Turkey	MSLQ
3	Cabedo & Masset Llaudes (2019)	0.04	92	4.17		College	Spanyol	custom
26	Fadlelmula et al. (2013) Study 3	-0.05	1019			JHS	Turkey	MSLQ
25	Fadlelmula et al. (2013) Study 2	-0.06	1019			JHS	Turkey	MSLQ
33	Tian et al. (2018) Study 6	-0.14	569			SHS	Taiwan	MKMQ
31	Tian et al. (2018) Study 4	-0.2	569			SHS	Taiwan	MKMQ
32	Tian et al. (2018) Study 5	-0.2	569			SHS	Taiwan	MKMQ

Note: ES = Elementary School; JHS = Junior High School; SHS = Senior High School

This study generally analyzes studies that produce data on the relationship between self-assessment and mathematics achievement at various levels of education. In this study, self-assessment was carried out through various existing instruments (MSLQ, MKMQ, Jr. MAI, MSA) and those developed by researchers with categories related to self-assessment (Rosário et al., 2013). These instruments often relate to students' self-regulated and metacognitive learning (Youde, 2019), while the mathematics learning achievement referred to in this study includes learning outcomes of mathematics material (in various domains) in specific periods (Ahmed et al., 2013) or special mathematics tests (namely MAT) conducted in a research by Özsoy (2011).

Table 2. Heterogeneity Residual Test

Fixed and Random Effects			
	Q	**df**	**P**
Omnibus test of Model Coefficients	77.690	1	< .001
Test of Residual Heterogeneity	942.670	42	< .001

Note. p -values are approximate.
Note. The model was estimated using the Restricted ML method.

From the various studies collected and selected, each effect size and standard error effect size were analyzed and then processed. The heterogeneity test results obtained are presented in Table 3.

Table 3. Heterogeneity Residual Estimation

	Estimates
I^2 (%)	95.344

Table 2 shows that the 43 study effect sizes analyzed were heterogeneous with a p <0.05 (Q=942.670; p<0.001). In addition, Arlinwibowo et al. (2021) also state that one method for testing heterogeneity is I^2, where I^2 illustrates the proportion of the variation in the summary effect size on a scale of 0% to 100%. Table 3 shows the value of I^2 = 95.344% > 25%, so it can be said that there is heterogeneity. Thus, the random effect model is more suitable for estimating the average effect size of the 43 studies analyzed. The analysis results also indicate the potential to investigate moderator variables influencing the relationship between self-assessment and students' mathematics achievement.

Table 4. Summary Effect/Average Effect Size

	Estimate	**Standard Error**	**z**	**p**	**95% Confidence Interval**	
					Lower	**Upper**
intercept	0.295	0.033	8.814	< .001	0.229	0.360

Note. Wald test.

Table 5. File Drawer Analysis

	Fail-safe N	**Target Significance**	**Observed Significance**
Rosenthal (1979)	19623.000	0.050	< .001

The results of the analysis using the random effect model show (see Table 4) that there is a significant positive correlation between self-assessment and students' mathematics achievement (z=8.814; p<0.001; 95% CI [0.229; 0.360]). The effect of self-assessment on students' mathematics learning achievement is in the low category, with r = 0.295 (Li, 2022). The overall description of the effect size and summary effect size of this study is illustrated in the forest plot in Figure 2. It shows that the effect size of the selected studies ranges from -0.20 to 1.09.

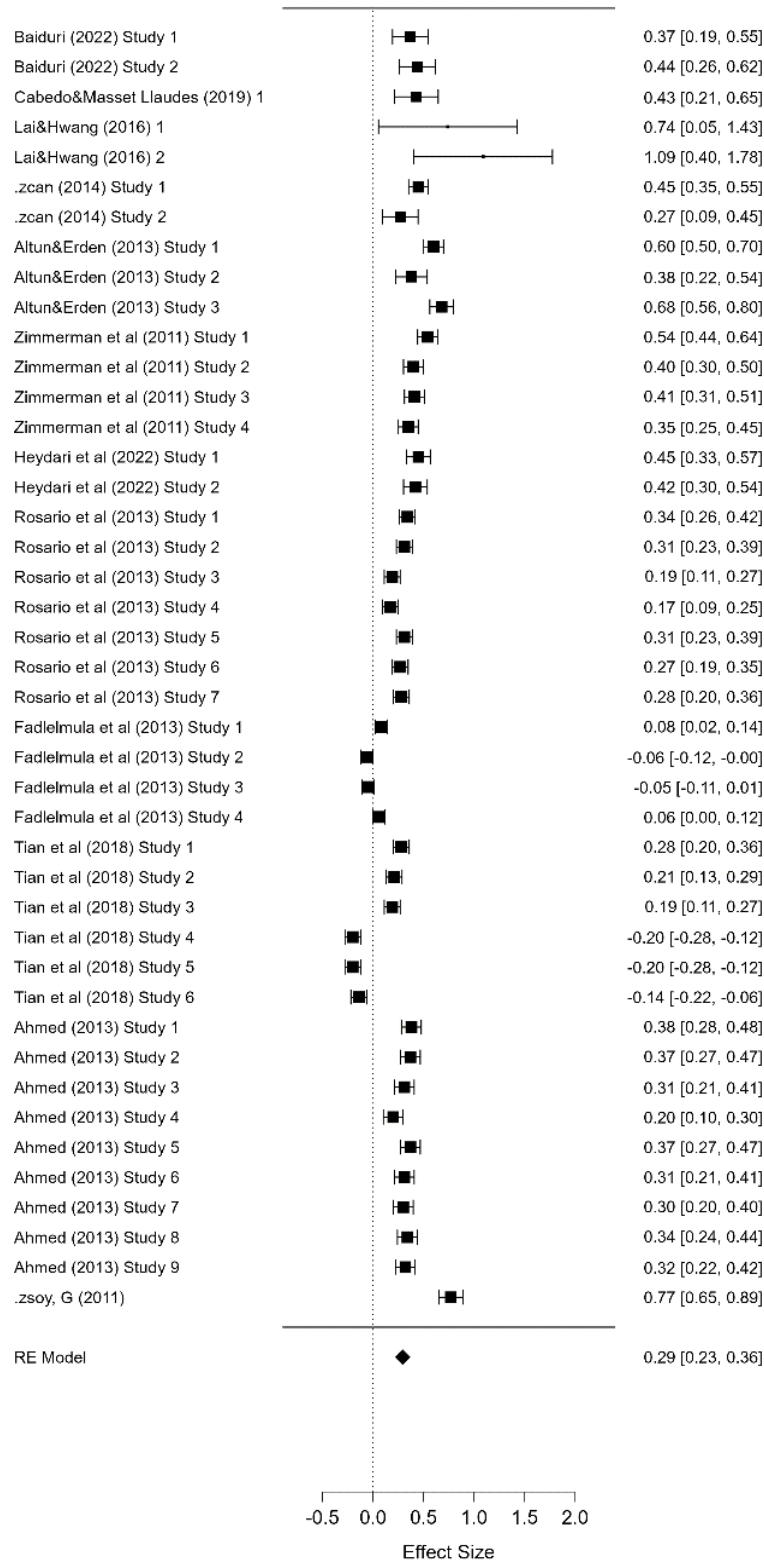|	
Baiduri (2022) Study 1	0.37 [0.19, 0.55]
Baiduri (2022) Study 2	0.44 [0.26, 0.62]
Cabedo&Masset Llaudes (2019) 1	0.43 [0.21, 0.65]
Lai&Hwang (2016) 1	0.74 [0.05, 1.43]
Lai&Hwang (2016) 2	1.09 [0.40, 1.78]
.zcan (2014) Study 1	0.45 [0.35, 0.55]
.zcan (2014) Study 2	0.27 [0.09, 0.45]
Altun&Erden (2013) Study 1	0.60 [0.50, 0.70]
Altun&Erden (2013) Study 2	0.38 [0.22, 0.54]
Altun&Erden (2013) Study 3	0.68 [0.56, 0.80]
Zimmerman et al (2011) Study 1	0.54 [0.44, 0.64]
Zimmerman et al (2011) Study 2	0.40 [0.30, 0.50]
Zimmerman et al (2011) Study 3	0.41 [0.31, 0.51]
Zimmerman et al (2011) Study 4	0.35 [0.25, 0.45]
Heydari et al (2022) Study 1	0.45 [0.33, 0.57]
Heydari et al (2022) Study 2	0.42 [0.30, 0.54]
Rosario et al (2013) Study 1	0.34 [0.26, 0.42]
Rosario et al (2013) Study 2	0.31 [0.23, 0.39]
Rosario et al (2013) Study 3	0.19 [0.11, 0.27]
Rosario et al (2013) Study 4	0.17 [0.09, 0.25]
Rosario et al (2013) Study 5	0.31 [0.23, 0.39]
Rosario et al (2013) Study 6	0.27 [0.19, 0.35]
Rosario et al (2013) Study 7	0.28 [0.20, 0.36]
Fadlelmula et al (2013) Study 1	0.08 [0.02, 0.14]
Fadlelmula et al (2013) Study 2	-0.06 [-0.12, -0.00]
Fadlelmula et al (2013) Study 3	-0.05 [-0.11, 0.01]
Fadlelmula et al (2013) Study 4	0.06 [0.00, 0.12]
Tian et al (2018) Study 1	0.28 [0.20, 0.36]
Tian et al (2018) Study 2	0.21 [0.13, 0.29]
Tian et al (2018) Study 3	0.19 [0.11, 0.27]
Tian et al (2018) Study 4	-0.20 [-0.28, -0.12]
Tian et al (2018) Study 5	-0.20 [-0.28, -0.12]
Tian et al (2018) Study 6	-0.14 [-0.22, -0.06]
Ahmed (2013) Study 1	0.38 [0.28, 0.48]
Ahmed (2013) Study 2	0.37 [0.27, 0.47]
Ahmed (2013) Study 3	0.31 [0.21, 0.41]
Ahmed (2013) Study 4	0.20 [0.10, 0.30]
Ahmed (2013) Study 5	0.37 [0.27, 0.47]
Ahmed (2013) Study 6	0.31 [0.21, 0.41]
Ahmed (2013) Study 7	0.30 [0.20, 0.40]
Ahmed (2013) Study 8	0.34 [0.24, 0.44]
Ahmed (2013) Study 9	0.32 [0.22, 0.42]
.zsoy, G (2011)	0.77 [0.65, 0.89]
RE Model	0.29 [0.23, 0.36]

Figure 2. Forest Plot

Publication bias in this meta-analysis study uses the theory by Rosenthal (1979), namely by using file drawer analysis. Because K=43, so 5K+10 = 5(43)+10 =225. The fail-safe N value obtained was 19623 (see Table 5), with a significance level of 0.05 and p <0.001. Because the value of fail-Safe N > 5K+10, it can be concluded that there is no problem of publication bias in the meta-analysis study (see Table 5).

The moderator variable analysis was carried out by determining each independent sample's mean weighted correlation. All 43 correlations were generated, contributing to the overall correlation between self-assessment and mathematics learning achievement. The random effect model's mean weighted correlation is significant (Qw = 586.76 ; p <0.001). The education level variable found significant differences in the average effect sizes at the ES, JHS, SHS, and colleges (Qb=355.91; p=0.00). However, based on the follow-up test analysis (post hoc), the difference is only the ES-JHS, ES-SHS, and college-SHS levels (see Figure 3).
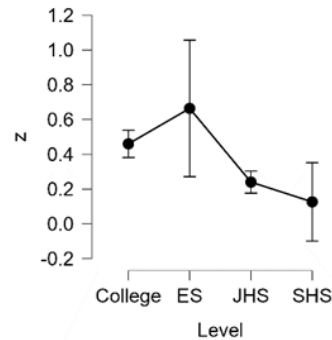
Figure 3. Education Level Descriptive Plot

As for regional variables, as shown in Figure 4, it was found that there was a significant difference in the average effect size in the eastern and western regions (Qb=202.548; p=0.00). However, based on the follow-up test analysis (post hoc), it turned out that there was no significant difference between eastern and western regions.
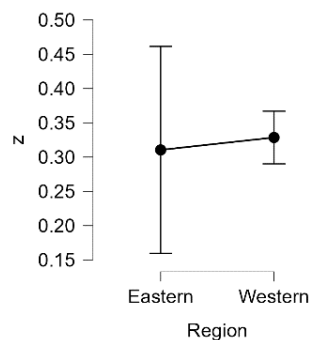

Figure 4. Research Region Descriptive Plot

Discussion

A well-designed self-assessment can support improving student learning performance (Brown & Harris, 2013). It must also be linked with instruction so students learn and perceive they are making progress (Schunk, 1996). Assessment practices that are tightly integrated into daily instruction will focus teacher attention on the objectives to be measured and provide teachers with more useful information in order to improve students' performance in mathematics (Ross et al., 2002). Through self-assessment, students can do meta-tasks to develop their meta-competencies (self-efficacy, self-confidence, motivation). They can encourage themselves to be able to study mathematics to a higher level of understanding (Beumann & Wegner, 2018). The advantages of self-assessment in supporting students' learning achievement are closely related to the relationship between self-assessment and self-regulated learning (Andrade, 2019; Yan et al., 2021). As previously found, self-regulated learning intervention had a good impact on at-risk urban technical college students. Self-regulated learning interventions positively affected the career's path of at-risk urban technical college students based on mathematics per-

formance benchmark analysis. However, there were 36% of the SRL group still did not meet the benchmark, so future research is needed to reveal the truth in this important group of students (Zimmerman, et.al., 2011).

The most robust relationship between self-assessment on mathematics learning achievement is shown in the self-regulated approach to mathematics learning achievement with a value of r = 0.80. This finding is in line with meta-analytical study by Panadero et al. (2017) which found a positive effect on students' SRL strategies and self-efficacy with self-assessment interventions. This indicates the need to use self-assessment within the framework of a self-regulated approach in the classroom to encourage students' mathematics learning achievement.

Schunk (1996) states that students who feel good about their self-evaluations are driven to keep working hard and feel effective about their learning. If students feel they are capable of achieving but that their current technique is unsuccessful, low self-judgments of progress and negative self-reactions will not always impair self-efficacy and drive. These students may change their self-regulatory processes by exerting more effort, persevering longer, choosing an approach they think is superior, or asking their classmates and instructors for assistance. Those students can be easily stimulated when learning in the scenario of self-regulation. In this situation, those teachers who plan to conduct in-class activities could care more for the lower self-regulated students. The lower self-regulated students need serious assistance to be able to improve their learning before class starts so that they will be better prepared to participate in learning in class (Lai & Hwang, 2016). Hopefully, their learning performance will improve, and they could be more engaged in the mathematics class activity. This engagement is needed in self-regulated learning approach to raise their mathematics learning achievement.

Meanwhile, the lowest r-value (r = -0.2) shows a negative correlation between metacognitive knowledge (MK) of self-difficulty/lack of fluency on mathematics achievement. This negative correlation indicates that the measurement of MK of self (difficulty/lack of fluency) and MK of tasks in MKMQ. The instrument may have limitations (Fadlelmula et al., 2014) or may not be appropriate for student participants who come from China because of the different subjectivity of 'difficulties' for students. Students in China are used to getting drills on difficult questions. In the context of MKMQ, it is classified as complex, but for Chinese students, it is easy (Tian et al., 2018). In other words, there is a need for a more thorough review of the related instrument items for further measures in the context of students who are different from the origin of the MKMQ instrument (Europe). Besides, Fadlelmula et al. (2014) also found that assessing students' learning strategies through self-report instruments, such as MSLQ, may have limitations for interpreting their actual implementation of these strategies. However, in the follow-up test of the moderator analysis, there was no significant difference in self-assessment and mathematics achievement in the two regions (eastern and western). Thus, there is a need for research related to the effects of self-assessment instruments adapted from the west in the context of eastern culture and self-assessment instruments adapted from the west originally on mathematics learning achievement.

A low correlation between self-assessment and students' mathematics learning achievement founded in the form of "metacognitive student-form" in the context of elementary school self-assessment. Students' age must be one of the affecting factors due this result besides the difficulty applying on-line measurement instruments like *think aloud protocols* or *interviews in crowded classrooms* which is common in Turkey (Özcan, 2014). However, Özcan (2014) also found that there is a strong correlation shown by the relationship between "metacognitive teacher-form" and students' mathematics achievement (r = .78). He adapted the instrument which is developed by Desoete (2007) namely Teacher Evaluation Form of Students' Metacognitive Abilities in Mathematics which is a mathematics domain specific measurement instrument.

Study by Özsoy (2011) also used an adapted version of the late Desoete's instrument namely MSA (Metacognitive Skills and Knowledge Assessment) (Desoete et al., 2001). The instrument's inventory was adapted to Turkish from the previous study in 2009 (Özsoy & Ataman,

2009; Özsoy et al., 2009). He found a high-level positive and reasonable relation between mathematics achievement and metacognitive skills (r = .648). In his research, the moderator's analysis also showed significant differences in self-assessment and achievement in learning mathematics at the elementary school level. Therefore, students at elementary school age need to be combined with self-assessment from teachers, considering that self-assessment skills at elementary school age still need to be developed (Nitko & Brookhart, 2014).

CONCLUSION

The selection of the random model in this meta-analysis is appropriate because the characteristics of the sample show homogeneous data and confirmation of an I2 value of more than 25%. This meta-analysis research shows a positive correlation of 0.29 between self-assessment and mathematics learning achievement. The meta-analysis results are free from publication bias, so the information obtained from the total effect size can be trusted. Even so, self-assessment must be designed properly to provide the optimal effect. This instrument design must also be ensured for its adjustment to the specific students' context. Then, this instrument design must meet the approach to learning mathematics in the classroom that supports the sustainability of the effects of the self-assessment for students. In this self-assessment context, self-regulated learning is one approach that correlates strongly with mathematics learning achievement. In addition, in the implementation of self-assessment, it is necessary to consider the students' age. For students at elementary school age, for example, it needs to be balanced by an assessment from the teacher or the need for parental assistance in its implementation.

REFERENCES

Ahmed, W., van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in Mathematics: A growth curve analysis. *Journal of Educational Psychology, 105*(1), 150–161. https://doi.org/10.1037/a0030160

Altun, S., & Elden, M. (2013). Self-regulation based learning strategies and self-efficacy perceptions as predictors of male and female students' mathematics achievement. *Procedia - Social and Behavioral Sciences*, *106*, 2354–2364. https://doi.org/10.1016/j.sbspro.2013.12.270

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*(August), 1–13. https://doi.org/10.3389/feduc.2019.00087

Arlinwibowo, J., Retnawati, H., & Kartowagiran, B. (2021). Item response theory utilization for developing the student collaboration ability assessment scale in STEM classes. *Ingenierie des Systemes d'Information*, *26*(4), 409–415. https://doi.org/10.18280/ISI.260409

Arlinwibowo, J., Retnawati, H., & Kartowagiran, B. (2022). The impact of ICT utilization to improve the learning outcome: A meta-analysis. *International Journal of Evaluation and Research in Education*, *11*(2), 522–531. https://doi.org/10.11591/ijere.v11i2.22112

Baiduri, B. (2022). Effect of self and peer assessments on mathematics learning achievement. *Al-Jabar: Jurnal Pendidikan Matematika*, *13*(1), 13–21. https://doi.org/10.24042/ajpm.v13i1.10731

Beumann, S., & Wegner, S. A. (2018). An outlook on self-assessment of homework assignments in higher mathematics education. *International Journal of STEM Education*, *5*(1), 1–7. https://doi.org/10.1186/s40594-018-0146-z

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367-393). SAGE. https://doi.org/10.4135/9781452218649.n21

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontiers of Learning Research, 2*(1), 22-30. https://doi.org/10.14786/flr.v2i1.24

Cabedo, J. D., & Maset-llaudes, A. (2019). How a formative self-assessment program positively influenced examination performance in financial mathematics. *Innovations in Education and Teaching International*, *57*(6), 680-690. https://doi.org/10.1080/14703297.2019.1647267

Castro-Alonso, J. C., Wong, M., Adesope, O. O., Ayres, P., & Paas, F. (2019). Gender imbalance in instructional dynamic versus static visualizations: A meta-analysis. *Educational Psychology Review, 31,* 361–387. https://doi.org/10.1007/s10648-019-09469-1

Cooper, H. M., Valentine, J. C., & Hedges, L. V. (2009). *Handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation. https://www.jstor.org/stable/10.7758/9781610441384

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Desoete, A., Roeyers, H., & De Clercq, A. (2001). Dynamic assessment of metacognitive skills in young children with mathematics-learning disabilities. In J. Carlson (Ed.), *Potential assessment and cognitive training: Actual research and perspectives in theory building and methodology*. JAI Press Inc./Elseiver.

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology*, 5(3), 705-730.

Fadlelmula, F. K., Cakiroglu, E., & Sungur, S. (2014). Developing a structural model on the relationship among motivational beliefs, self-regulated learning strategies, and achievement in Mathematics. *International Journal of Science and Mathematics Education, 13*, 1355–1375. https://doi.org/10.1007/s10763-013-9499-4

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educational Psychology Review, 34,* 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Lai, C., & Hwang, G. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education*, *100 September*, 126–140. https://doi.org/10.1016/j.compedu.2016.05.006

Li, Y. (2022). Statistical power analysis for the behavioral sciences. In B. B. Frey (Eds.), *The SAGE encyclopedia of research design* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4135/9781071812082.n600

Martin-Martin, A., Orduna-Malea, E., Harzing, A. W., & López-Cózar, E. D. (2017). Can we use Google Scholar to identify highly-cited documents? *Journal of Informetrics*, *11*(1), 152–163. https://doi.org/10.1016/j.joi.2016.11.008

McMillan, J. H. (2013). *Classroom assessment.* In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment,* pp. 1-3. SAGE.

Ministry of Education and Culture. (2016). *Panduan penilaian oleh pendidik dan satuan pendidikan untuk sekolah menengah pertama.* Kementerian Pendidikan dan Kebudayaan. Retreived from http://ditpJHS.kemdikbud.go.id

Ministry of Education, Culture, Research, and Technology. (2021). *Panduan pembelajaran dan asesmen (Jenjang pendidikan dasar dan menengah).* Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. https://kurikulum.kemdikbud.go.id/wp-content/uploads/2022/06/Panduan-Pembelajarn-dan-Asesmen.pdf

Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (6th ed.). Pearson Education, Inc.

Özsoy, G. & Ataman, A. (2009). The effect of metacognitive strategy training on problem solving achievement. *International Electronic Journal of Elementary Education*, 1(2), 67–82.

Özsoy, G., Memis, A., & Temur, T. (2009). Metacognition, study habits, and attitudes. *International Electronic Journal of Elementary Education*, 2(1), 154–166.

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. *Asia Pacific Education Review, 12*, 227–235. https://doi.org/10.1007/s12564-010-9129-6

Özcan, Z. Ç. (2014). Assessment of metacognition in Mathematics: Which one of two methods is a better predictor of Mathematics achievement?. *International Online Journal of Educational Sciences, 6*(1), 49–57. https://iojes.net/?mod=makale_tr_ozet&makale_id=41068

Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy, and Practice*, *21*(4), 365–383. https://doi.org/10.1080/0969594X.2014.919247

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

Popelka, E. (2015). *Improving the accuracy of middle school students' self-assessment, peer assessment, and mathematics achievement.* [Doctoral dissertation, University of Louisville]. https://doi.org/10.18297/etd/2324

Retnawati, H., Apino, E., Kartianom, K., Djidu, H., & Anazifa, R. D. (2018). *Pengantar analisis meta.* Parama Publishing.

Rosário, P., Núñez, J. C., Valle, A., & González-Pienda, J. (2013). Grade level, study time, and grade retention and their effects on motivation, self-regulated learning strategies, and mathematics achievement: A structural equation model. *European Journal of Psychology of Education, 28*, 1311–1331. https://doi.org/10.1007/s10212-012-0167-9

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem- solving achievement. *International Journal of*

Phytoremediation, *21*(1), 43–58. https://doi.org/10.1207/S15326977EA0801_03

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, *33*(2), 359–382. https://doi.org/10.3102/00028312033002359

Setiadi, H. (2016). Pelaksanaan penilaian pada Kurikulum 2013. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *20*(2), 166–178. https://doi.org/10.21831/pep.v20i2.7173

Tian, Y., Fang, Y., Li, J., Wang, L., & Greenshaw, A. J. (2018). the effect of metacognitive knowledge on Mathematics performance in self-regulated learning framework — Multiple mediation of self-efficacy and motivation. *Front. Psychol.*, *9*, 2518. https://doi.org/10.3389/fpsyg.2018.02518

Yan, Z. (2016). The self-assessment practices of Hong Kong secondary students: Findings with a new instrument. *Journal of Applied Measurement*, *17*(3), 335–353.

Yan, Z., Wang, X., Boud, D., & Lao, H. (2021). The effect of self-assessment on academic performance and the role of explicitness: A meta-analysis. *Assessment and Evaluation in Higher Education*, *48*(1), 1–15. https://doi.org/10.1080/02602938.2021.2012644

Youde, J. J. (2019). *Meta-analysis of the effects of reflective self-assessment on academic achievement in primary and secondary populations*. [Doctoral dissertation, Seattle Pacific University]. https://digitalcommons.spu.edu/soe_etd/48/

Zare, A., Heydari, H., Davoodi, H. & Kia, M. M. (2022). Modeling the effect of self-evaluation on the progress of second year high school students in mathematics with the mediation of self-regulation and mathematical self-efficacy. *Journal of School Psychology and Institutions*, *11*(2), 62-70. https://dx.doi.org/10.22098/jsp.2022.1717

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling*, *53*(1), 141–160. https://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2011_20110328/07_Zimmermann.pdf