



A Reliability Generalization Meta-Analysis of Mathematics Self Efficacy- Revised (MSES-R) Scale*

Hatice ÇETİN**, İbrahim ÇETİN***

Article Information	ABSTRACT
Received: 28.05.2025	The Mathematics Self-Efficacy-Revised (MSES-R) scale (Kranzler & Pajares, 1997) has been widely used to measure students' mathematics self-efficacy. This study aims to determine the common reliability coefficient using reliability generalization, a type of meta-analysis. The research process was organized using PRISMA guidelines. The average consistency reliability was calculated as 0.901 with a 95% CI [0.864; 0.939] using a random effect model. The studies showed heterogeneity [Q (6) = 219.788; $p < .01$; $I^2 = 96.99\%$]. The meta-analysis predicts that certain moderator variables significantly affect the scale's reliability. The reliability generalization analysis of the MSES-R scale reveals the different properties of the MSES-R inventory using a mixed effects model. It was found that sample type and department context significantly affect the Cronbach's alpha coefficient of the MSES-R scale ($p < 0.01$), while the study location variable does not significantly affect the common coefficient ($p > 0.01$).
Accepted: 14.01.2026	
Online First: 27.04.2026	
Published: 30.04.2026	
doi: 10.16986/hunefd.1708169	Keywords: Cronbach's alpha coefficient, mathematics self-efficacy, meta-analysis, reliability generalization
	Article Type: Research Article

Citation Information: Çetin, H., & Çetin, İ. (2026). A Reliability Generalization Meta-Analysis of Mathematics Self Efficacy- Revised (MSES-R) Scale. *Hacettepe University Journal of Education*, 41(2), 394-403. <https://doi.org/10.16986/hunefd.1708169>

1. INTRODUCTION

Educational psychologists and mathematics educators have long sought variables that can enhance students' cognitive and affective outcomes in mathematics learning, particularly those that foster academic achievement (Chief Examiner Report, 2005, 2006, 2007). Among these variables, self-efficacy beliefs have received increasing attention. Self-efficacy is defined as individuals' evaluations of their competence, skills, and capacities and their ability to translate these into behaviors (Schunk, 2009). According to Bandura (1986), self-efficacy reflects individuals' judgments about their capacity to perform a specific task, and it is considered a key determinant of motivation and behavior (Pajares, 1992). Individuals with strong self-efficacy beliefs are more motivated to cope with difficulties and exert greater effort to overcome challenges (Bandura, 1994), whereas those with weak self-efficacy beliefs tend to give up quickly when facing obstacles (Pajares, 1996).

In mathematics education, affective factors are as important as cognitive skills (McLeod, 1992). One of the most influential affective factors associated with mathematics achievement is mathematics self-efficacy (Usher, 2009). It is generally defined as students' judgments about their ability to successfully complete mathematical tasks (Hackett & Betz, 1989; Pajares & Kranzler, 1995). Students' confidence in their mathematical abilities enables them to engage actively in learning by applying different strategies, methods, and techniques (Alderman, 1999). Pajares and Kranzler (1995) even argued that the predictive power of self-efficacy beliefs for mathematics performance is as strong as that of general mental ability. Indeed, numerous studies have consistently reported a significant relationship between mathematics self-efficacy and achievement (Hackett & Betz, 1989; Moore, 2005; Pajares & Graham, 1999; Pajares & Miller, 1995; Usher, 2009). Collectively, these findings emphasize the necessity of developing students' mathematics self-efficacy beliefs to improve their achievement.

Accurate measurement of affective constructs such as mathematics self-efficacy is therefore crucial for evaluating students' motivation in learning mathematics. The reliability of measurement instruments directly affects educational interventions and instructional decisions. Among the available instruments, the Mathematics Self-Efficacy Scale-Revised (MSES-R; Kranzler & Pajares, 1997) stands out as one of the most frequently validated and widely applied tools to assess mathematics self-efficacy.

* This study was presented at a scientific event. Çetin, H., & Çetin, İ. (2023). Reliability generalization meta-analysis of the Revised Mathematics Self-Efficacy Scale (MSES-R). [Oral Presentation]. International Education Congress 2023. Since the study did not involve human practices, there was no need to obtain an ethics committee approval.

** Assoc. Prof. Dr., Necmettin Erbakan University, Ahmet Keleşoğlu Faculty of Education, Department of Primary Education, Division of Classroom Education, Konya-TÜRKİYE. e-mail: haticecetin@erbakan.edu.tr (ORCID: 0000-0003-0686-8049)

*** Assoc. Prof. Dr., Necmettin Erbakan University, Ahmet Keleşoğlu Faculty of Education, Department of Math and Science Education, Division of Math Education, Konya-TÜRKİYE. e-mail: ibrahimcetin@erbakan.edu.tr (ORCID: 0000-0003-4807-3295)

It consists of three subscales—tasks, courses, and problems—that capture students' confidence in applying mathematics to real-life tasks, succeeding in mathematics courses, and solving novel problems. Specifically, the tasks subscale measures students' confidence in applying mathematics to everyday situations such as calculating recipe quantities; the courses subscale reflects their beliefs about succeeding in mathematics subjects like geometry or algebra; and the problems subscale evaluates their confidence in solving novel mathematical problems.

Compared to other instruments such as the Mathematics Self-Efficacy Questionnaire (MSEQ; Betz & Hackett, 1983) or the Self-Efficacy for Learning Mathematics Scale (SELM; Bandura, 1997), the MSES-R provides a broader and more differentiated measurement framework that captures both academic and applied aspects of mathematical competence (Pajares & Kranzler, 1995). Its tripartite structure allows researchers to distinguish between students' perceived competence in formal coursework and their confidence in problem-solving or real-life applications, offering a more holistic view of mathematics self-efficacy. Furthermore, the MSES-R has been extensively validated and adapted across diverse populations and cultural contexts, resulting in a relatively large and methodologically diverse body of literature (e.g., Finney & Schraw, 2003; Guardia-Olmos et al., 2012; Blake & Masten, 2015). This makes it particularly suitable for a reliability generalization (RG) meta-analysis, as the accumulation of independent reliability coefficients provides a strong empirical basis for evaluating its internal consistency across contexts. Therefore, the MSES-R was selected for examination in this study.

Despite its widespread application, no RG meta-analysis has yet synthesized this evidence. Therefore, this study not only addresses the lack of RG research on MSES-R but also establishes its theoretical and empirical superiority over other mathematics self-efficacy instruments. By synthesizing reliability coefficients from diverse studies, it aims to provide a comprehensive understanding of the MSES-R's measurement precision and to strengthen its role as a benchmark tool in mathematics self-efficacy research.

Prior RG studies in mathematics education have addressed constructs including motivation (Taylor, 2012; Holland et al., 2018), learning (Mji & Alkhateeb, 2005; Henson & Hwang, 2002; Ross, Blackburn & Forbes, 2005), self-efficacy (Henson, Kogan & Vacha-Haase, 2001; Nilsson, Schmidt & Meek, 2002), thinking (Mahapoonyanont, Krahamwong, Kockhakornjarupong & Rachasong, 2010), self-esteem (Lane, White & Henson, 2002), anxiety (Capraro, Capraro & Henson, 2001; Barnes, Harp & Jung, 2002), self-awareness (Leach, Henson, Odom & Cagle, 2006), personality (Caruso, Witkiewitz, Belcourt-Ditloff & Gotlieb, 2001; Vacha-Haase, Kogan, Tan & Woodall, 2001; Capraro & Capraro, 2002), well-being (Crouch, 2016), and behavior (Hess, McNab & Basoğlu, 2014; Kilgus, Eklund, Maggin, Taylor & Allen, 2018). In addition, the Career Decision Self-Efficacy Scale (CDSE) has been the subject of an RG study focusing on career decision-making (Nilsson, Schmidt & Meek, 2002). Reliability studies have also been conducted on self-efficacy as a sub-dimension of motivational learning strategies (Taylor, 2012; Holland et al., 2018). However, no RG study has focused exclusively on a scale designed to measure mathematics self-efficacy. Addressing this gap, the present study conducts a RG of the widely used problem dimension of the Mathematics Self-Efficacy Scale-Revised (MSES-R).

2. METHODOLOGY

In this section, the type of the study, the data collection process, the process of determining the criteria, the validity and reliability of the measurement results related to the coding form, and the data analysis are mentioned.

2.1. Research Type

To achieve this goal, we conducted a systematic review and analysis following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PRISMA is preferred for managing and reporting literature reviews. The four steps of PRISMA are as follows:

1. Identifying research literature from database searches
2. Screening articles using inclusion and exclusion criteria
3. Assessing full-text articles for eligibility
4. Analyzing and reporting the final articles included in the review.

In the initial stage, studies were identified through searches in Web of Science, Scopus, and Google Scholar databases to ensure comprehensive coverage of the literature. However, after applying the inclusion and exclusion criteria, only the studies indexed in the Web of Science database met all requirements, and thus the final analysis was conducted on these seven studies. The study publication period for this review was from 1997 to December 2023 since Kranzler & Pajares published MSES-R in 1997, with no study conducted in 2024. First, studies that cited the relevant article (Kranzler & Pajares, 1997) were identified in the specified databases. The study was limited to quantitative methods

2.2. Meta-Analysis of Reliability and Reliability

Reliability is related to the consistency and reproducibility of the scores of a measurement tool administered at a certain time and under specific conditions (Onwuegbuzie & Larry, 2000; Taylor, 2012). Therefore, the reliability of a measurement tool is affected by the sample size and sample characteristics (Ashley, Williams & Young, 2021). Vacha-Haase and Thompson (2011)

stated that reporting reliability coefficients is essential. If the reliability coefficient is not reported, this omission jeopardizes the study's reproducibility under similar conditions (Cousin & Henson, 2000).

As a result of the importance and incentives for reporting reliability coefficients, a new method was developed to investigate reliability coefficients and reach a common reliability coefficient (Ashley, Williams & Young, 2021). This new approach, described by Vacha-Hasse (1998) and known as RG or RG meta-analysis, is a tool developed to identify the source of differences among coefficients and present reliability coefficients holistically. RG aims to identify the source of measurement error in studies using the same instrument. It provides researchers with a method to assess score reliability in previous administrations of an instrument and investigate possible sources of variability in reliability estimates. Additionally, RG enables researchers to estimate expected reliability levels in their studies and provides options regarding different research designs concerning statistical significance and effect size (Henson & Thompson, 2002). Studies using the relevant measurement tool are data sources, and reliability coefficients are coded as dependent variables, while sample characteristics are coded as independent variables (possible predictors) (Cousin & Henson, 2000).

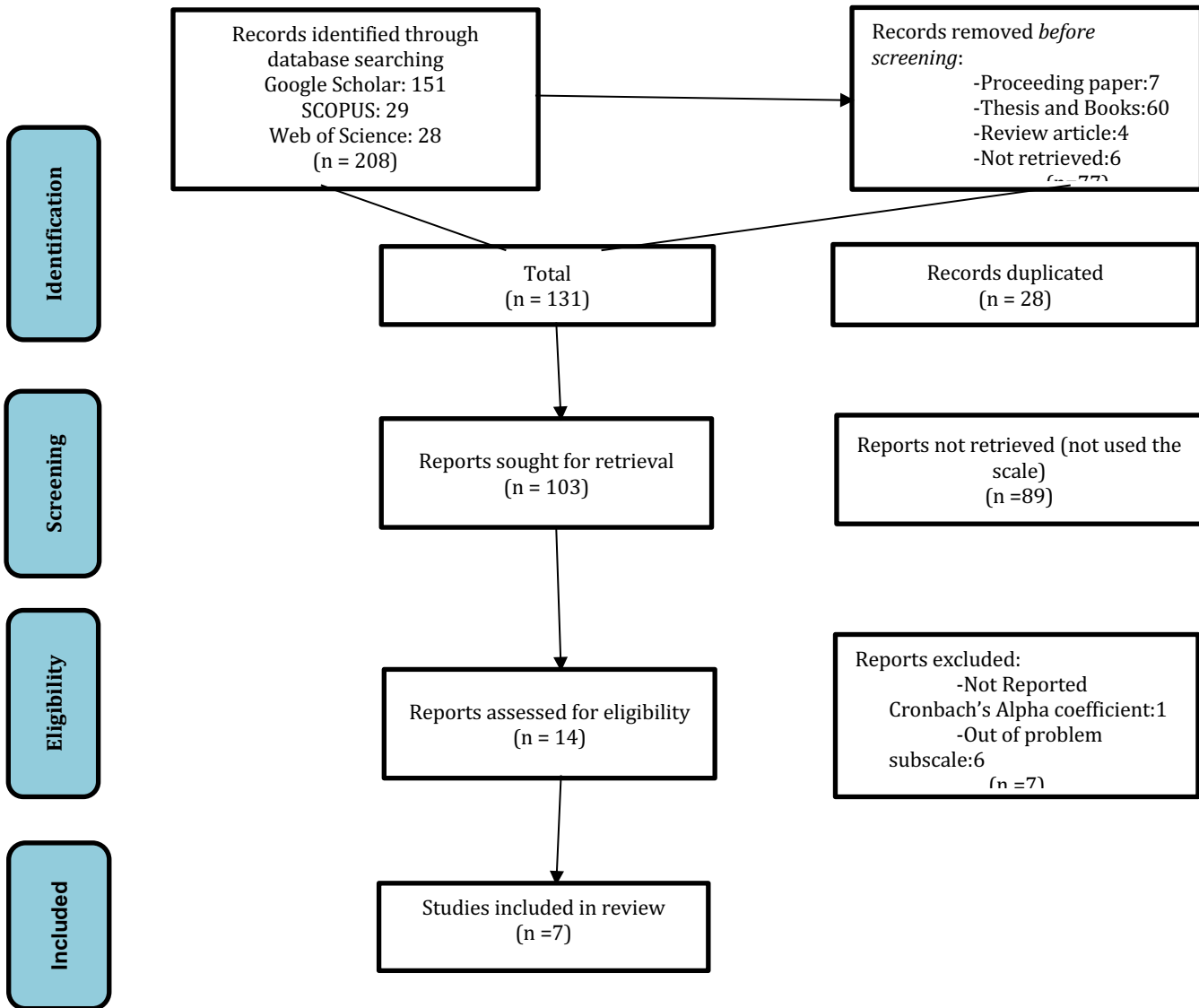


Figure 1. Study selection flow diagram

2.3. Data Collection Process

This study was initiated to identifying and reviewing research articles that cited the Mathematical Self-Efficacy Scale – Revised (MSES-R) developed by Kranzler and Pajares (1997), in order to systematically examine the empirical studies employing this scale and to include them in a RG meta-analysis.

As shown in Figure 1, electronic database searches were conducted using Google Scholar, Scopus, and Web of Science, yielding a total of 208 records, including 151 from Google Scholar, 29 from Scopus, and 28 from Web of Science. After removing proceeding

s ($n = 7$), theses and books ($n = 60$), review articles ($n = 4$), and studies that could not be retrieved ($n = 6$), seventy-seven records were excluded prior to screening. An additional twenty-eight duplicate records were also removed. The remaining studies were screened based on the predetermined inclusion and exclusion criteria.

The inclusion criteria required that studies had employed the Mathematical Self-Efficacy Scale (MSES-R) had reported Cronbach's alpha coefficient or another internal consistency reliability coefficient for the overall scale or its subscales, been published in peer-reviewed journals, and had provided sufficient statistical data for RG analysis. In contrast, studies were excluded if they were review papers, theses, books, or conference proceedings, did not employ the MSES-R, failed to report a Cronbach's alpha coefficient, or contained duplicated data.

Following this process, eighty-nine studies were excluded for not using the relevant scale, leaving fourteen studies for full-text review. Among these, one study was excluded for not reporting Cronbach's alpha coefficient, and six were excluded for reporting reliability estimates related to subscales outside the problem-solving dimension. Consequently, seven studies met all inclusion criteria and were retained for the final RG meta-analysis. Figure 1 presents the PRISMA flow diagram summarizing the overall study selection process.

2.4. Mathematics Self-Efficacy-Revised Scale

In mathematics education literature, numerous scales have been developed to measure mathematics self-efficacy. MSES-R is the most preferred instrument in mathematics education studies. MSES-R was initially developed by Kranzler & Pajares (1997) with 52 items. They later revised the scale, and calculated Cronbach's alpha coefficient of 0.901 during the original study. The scale, measured on a 5-point Likert scale, consists of three dimensions: tasks, courses, problems. Table 1 shows the items in the MSES-R.

Table 1.
Scale Items in the MSES-R Scale

Subscale	Item (example)
Tasks	Calculate recipe quantities for a dinner for 41 when the original recipe is for 12
Courses	Geometry
Problems	A baseball player got two hits in three at bats. This could be represented by $2/3$. Which decimal would most closely represent this?

2.5. Data Analysis

The 14 full-text articles were re-evaluated to confirm the presence of information about the reliability coefficients. One study that did not report Cronbach's alpha coefficient was excluded. Only studies with full-text availability in English were included. Finally, seven studies met all inclusion criteria and were analyzed using the RG meta-analytic approach.

A random-effects model was applied because variability in Cronbach's alpha coefficients was expected across studies due to differences in sample characteristics, populations, and research contexts (Borenstein et al., 2019; Hedges & Vevea, 1998). The between-study variance (τ^2) was estimated using the Restricted Maximum Likelihood (REML) method, which yields more stable and unbiased estimates compared to other estimators such as DerSimonian-Laird (Viechtbauer, 2005). Common heterogeneity statistics (Cochran's Q and I^2) were calculated to quantify the variability across studies.

Publication bias was assessed using Rosenthal's Fail-Safe N , supplemented by Egger's regression and Kendall's tau tests to ensure robustness. Additionally, a mixed-effects model was employed to examine whether categorical moderators (e.g., sample type, discipline, study location) significantly influenced the overall reliability estimates. All analyses were performed using Jamovi software (Version 2.3.13).

3. RESULTS

In this section of the study, a meta-analysis was conducted to calculate the overall effect size concerning the reliability of the Mathematics Self-Efficacy Scale, as measured by Cronbach's alpha coefficient. Additionally, moderator analyses were used to identify the sources of variation in the average Cronbach's alpha coefficients.

3.1. Reliability Findings and Sample Distribution

This section first presents the findings on publication bias, followed by the results of heterogeneity, effect size, and moderator analyses. The characteristics of the included articles are summarized in Table 2.

Table 2.
Mean of Cronbach' Alpha Coefficients, 95% Confidence Intervals, and Heterogeneity Tests for the Problem Dimension of MSES-R

Authors	Country-Continent	Sample Type	Discipline	n	Alpha
Guardia-Olmos et al. (2012)	Spain-Europe	University	Psychology	96	0.916
Cebollero et al. (2012)	Spain-Europe	University	Psychology	95	0.916
Finney & Schraw (2002)	USA-America	University	Psychology	138	0.92
Dobarro (2007)	Brazil-Europe	High School	General	213	0.9
Blake & Masten (2015)	Jamaica-USA	High School	General	780	0.97
Ayatola & Adedeji (2009)	Nigeria-Africa	High School	General	352	0.87
Briley (2012)	USA-America	Adults	Primary School Teachers	95	0.78

Table 2 shows the features of the analyzed studies regarding country-continent, sample type, discipline, number of participants, and Cronbach’s alpha coefficient. The studies were distributed as follows: three studies from Europe, three from the USA, and one from Africa. Most studies were conducted in high schools (n=1345), three studies were conducted with university students in psychology (n=329), and one study was conducted with adults (primary school teachers, n=95). The cumulative sample size consisted of 1769 participants. The “problem” dimension was included in the scale as part of the developed MSES-R scale. Cronbach’s alpha coefficients varied between 0.78 and 0.97.

3.2. Reliability and Heterogeneity

The Cronbach’s alpha coefficients reported for the overall scale across seven studies were determined to assess reliability. The total number of participants was 1769 (range n = 95 to n = 780) with M = 1196 and sd = 924. The statistics of the RG of the MSES-R scale are presented in Table 3.

Table 3. Mean of Cronbach’s Alpha Coefficients, 95% Confidence Intervals, and Heterogeneity Tests for the Problem Dimension of MSES-R

Scale	N	M	sd	CA	LL	UL	Q	I2
Problem	7	1196	924	0.901	0.864	0.939	219.788	96.99%

Note: n = Number of Studies; CA= Cronbach’s Alpha Coefficient LL = Lower Limit; UL = Upper Limit; Q = Cochran’s Heterogeneity Q Statistic; I2 = Heterogeneity Index; p < 0.01.

According to the RG of the random-effects meta-analysis, the cumulative Cronbach’s alpha coefficient is 0.901 (95% CI [0.864; 0.939]). This value was calculated using the Cronbach’s alpha coefficients reported in the studies. Heterogeneity was assessed using Cochran’s Q and I2 tests on the Cronbach’s alpha coefficients. The Cochran’s Q test yielded a significant result (df(6) = 219.788; p < .01), while the I2 test indicated a high level of heterogeneity (96.99%). This substantial heterogeneity suggests the need for further analysis. Heterogeneity among the calculated coefficients was analyzed by constructing a forest plot. Figure 2 presents a forest plot of Cronbach’s alpha coefficient for the problem dimension of the MSES-R scale.

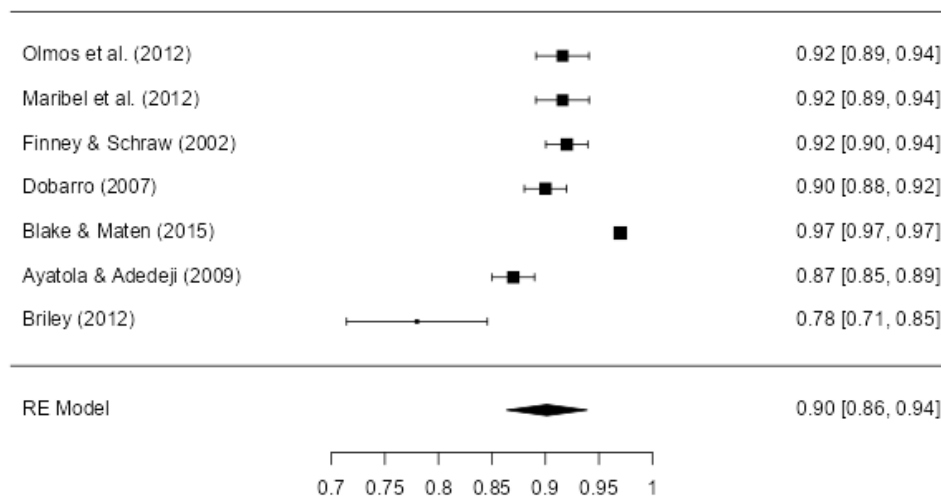


Figure 2. Forest plot for the RG of MSES-R Scale

3.3. Results Regarding Publication Bias

Within the scope of the study, before calculating the mean effect size for the meta-analysis, researchers attempted to detect any signs of publication bias in the studies reviewed during the analysis. The Fail-Safe N test, Egger's Regression test, and Kendall's Tau test were used to reveal any possible publication bias.

Table 4.

Publication Bias Assessment

Test	Value	P Value
Fail-Safe N	411857.000	< .001
Egger's Regression Intercept	-4.126	< .001
Kendall's Tau (τ)	-0.143	0.773

The Fail-Safe N test yielded a value of 411857.000 ($p < .001$). The intercept in Egger's Regression was calculated to be -4.126 ($p < .001$). An examination of the results of Kendall's Tau test ($\tau = -0.143$; $p = .773$) using the funnel plot concluded that there is no publication bias. Figure 2 presents the funnel plot based on the transformed Cronbach's alpha coefficient for publication bias. The results show that there is no publication bias.

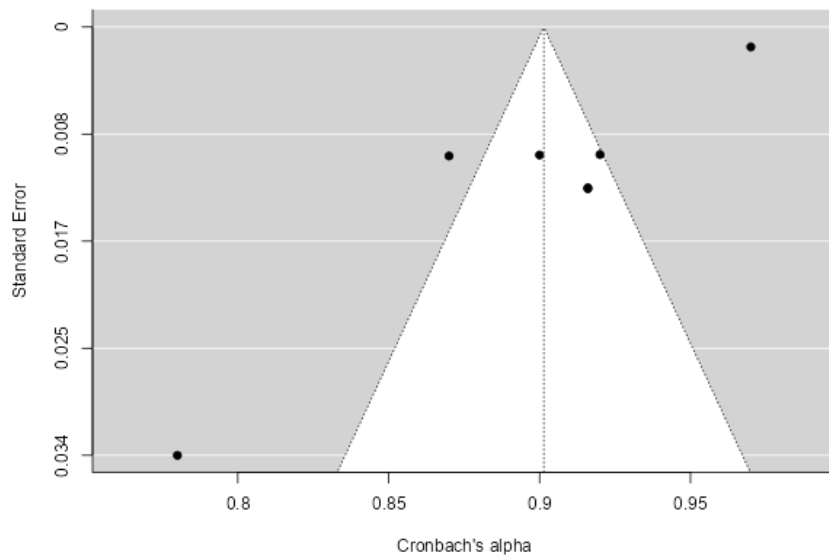


Figure 2. Funnel plot of Cronbach's coefficient for the MSES-R Scale

To address potential publication bias, a funnel plot was produced followed by Egger's test. Funnel plot asymmetry was suggested by Egger test, producing a p-value less than 0.001.

3.4. Results Related to Moderator Analyses

In RG research, there are different approaches to evaluate moderator variables and explain the variability of reliability coefficients. Some researchers propose various descriptive and inferential statistical strategies to explain the variability of reliability coefficients. Strategies for investigating variability include methods suggested by researchers like such as Vacha-Haase, Tani, and others (2001).

Moderator analyses were conducted using the mixed-effects model to investigate the reliability values of categorical variables. The categorical variables considered were sample type (university, high school, adults), study location (Europe, America, Africa), and department (psychology, high school, primary school teaching). The analysis found that school level significantly affects the discipline ($z = -0.0861$; $p < 0.01$; 95% CI [-0.063; 0.058]; $R^2 = 43.81\%$), as does the discipline ($z = -2.53$; $p < 0.01$; 95% CI [-0.244; -0.0031]; $R^2 = 43.81\%$). No significant effect was seen with the country-continent ($z = -2.25$; $p > 0.01$; 95% CI [-0.113; 0.089]; $R^2 = 0\%$).

This study examined the reliability of the Mathematics Self-Efficacy Scale-Revised (MSES-R) using the RG meta-analysis method. The primary aim was to determine the common reliability coefficient of the MSES-R scale. The findings contribute significantly to understanding the different effects on the reliability of this scale to and synthesizing various results in the literature.

The results indicate that the overall reliability of the MSES-R scale, calculated using the random effect model, is 0.901 (95% CI [0.864; 0.939]). However, these results showed substantial heterogeneity [$Q(6) = 219.788$; $p < 0.01$; $I^2 = 96.99\%$]. This high

level of heterogeneity underscores the importance of investigating the sources of measurement error in studies using the same instrument (Henson & Thompson, 2002).

The meta-analysis results predict that certain moderator variables significantly impact the reliability of the MSES-R scale. Specifically, sample type (university, high school, adults) and discipline (psychology, high school, primary school teaching) had significant effects on the overall estimate. This finding suggests that different sampling groups can influence the reliability of the scale, particularly given the high heterogeneity observed. Additionally, the results indicate that the country-continent variable does not significantly affect the overall reliability estimate. This suggests that the MSES-R scale may yield consistent reliability results across different geographical regions, supporting its global applicability.

These findings imply that factors such as sample type and discipline should be considered when interpreting results using the MSES-R scale. The reliability of a measurement tool is influenced by the type and number of the sample (Ashley, 2021). Therefore, reporting the common reliability coefficient of studies is recommended (Cousin & Henson, 2000).

This study provides further insight into the reliability of the MSES-R scale and emphasizes the importance of accurately measuring mathematics self-efficacy. Future research should explore the factors influencing the reliability of the MSES-R scale in greater detail and offer additional insights into the use of this scale. These findings can enhance the understanding of the tools used to assess and develop students' mathematics self-efficacy. Therefore, it is crucial to consider differences in student groups and disciplines when using the MSES-R scale.

3.5. Strengths and Limitations

This study provides a comprehensive and methodologically robust synthesis of the reliability of the Mathematics Self-Efficacy Scale-Revised (MSES-R) through a RG meta-analysis. By integrating results from seven studies conducted across three continents (Europe, America, and Africa) and diverse educational levels, the research offers valuable evidence of the scale's high internal consistency ($\alpha = 0.901$). The use of a random-effects model strengthens the reliability and accuracy of the findings. Moreover, the inclusion of moderator analyses, covering variables such as sample type, department, and geographical region adds explanatory depth by identifying contextual factors that influence reliability estimates. The results demonstrating the MSES-R's consistent reliability across different regions underscore the global applicability and psychometric robustness of the scale. Furthermore, the study followed PRISMA guidelines, ensuring transparency and replicability in the research process. Collectively, these strengths position the present study as a significant contribution to the literature on mathematics self-efficacy measurement and meta-analytic reliability research.

The cumulative Cronbach's alpha coefficient indicates a high level of internal consistency for the MSES-R scale across studies. However, the heterogeneity was substantial, suggesting that reliability estimates varied considerably depending on study characteristics. Moderator analyses (sample type and department) showed some differences but given the small number of studies ($k=7$), these findings should be interpreted with caution. Notably, the gender variable, commonly examined in RG studies, could not be included, as none of the available studies reported gender-specific coefficients.

4. DISCUSSION, CONCLUSION AND SUGGESTIONS

The MSES-R scale has been widely used at all levels of education across various countries, demonstrating good reliability. This study utilized the RG meta-analysis method to examine the reliability of the Mathematics Self-Efficacy Scale-Revised (MSES-R). The primary objective was to identify the common effect on the reliability that the MSES-R scale may have. The findings significantly contribute to analyzing various factors affecting the reliability of this scale and synthesizing various results from the literature.

The overall reliability of the MSES-R scale was calculated as 0.901 (95% CI [0.864; 0.939]) using the random effect model. However, these results showed significant heterogeneity [$Q(6) = 219.788$; $p < 0.01$; $I^2 = 96.99\%$], emphasizing the importance of investigating the sources of measurement error in studies using the same instrument (Henson & Thompson, 2002).

The study predicts that certain moderator variables significantly impact the reliability of the MSES-R scale. Specifically, sample type (university, high school, adults) and discipline (psychology, high school, primary school teaching) significantly affect the overall estimate. Different sampling groups can influence the reliability of the scale. Additionally, the continent variable does not significantly affect the overall reliability estimate, suggesting that the MSES-R scale may yield consistent reliability results across different geographical regions.

The high reliability of the MSES-R across diverse populations and cultural contexts reinforces its value as a benchmark instrument in both research and educational practice. Its consistent psychometric performance supports its use for diagnosing students' self-efficacy beliefs, designing interventions to strengthen mathematics confidence, and enabling valid cross-cultural and cross-level comparisons. These findings also highlight the importance of considering sample type and academic discipline when interpreting results obtained through the MSES-R scale, as variations in participant characteristics may influence reliability outcomes. Reporting the common reliability coefficient of studies using this instrument is therefore recommended to enhance comparability across research (Cousin & Henson, 2000).

Overall, this study not only contributes methodologically by synthesizing reliability evidence for the MSES-R but also reinforces its theoretical and practical significance as a key tool in understanding how students develop, experience, and apply mathematics self-efficacy. Future research should examine additional factors influencing the reliability of the MSES-R, such as cultural context, gender, and instructional environment, to further clarify its robustness. By doing so, researchers and educators can gain deeper insights into how mathematics self-efficacy can be more accurately assessed and effectively enhanced across diverse educational settings.

Research and Publication Ethics Statement

This study is a systematic literature review examining the reliability coefficients of the Mathematics Self-Efficacy Scale. Data were not collected from any group of students or teachers; data were obtained solely by accessing studies in the literature. Therefore, no application was made to the Ethics Commission for this study.

Contribution Rates of Authors to the Article

Researchers contributed equally at all stages of the research.

Statement of Interest

The authors have not declared any potential conflicts of interest regarding the research, authorship, or publication of this article.

5. REFERENCES

Studies included in the current RG meta-analysis are marked with an asterisk (*)

Ayotola, A., & Adedeji, T. (2009). The relationship between mathematics self-efficacy and achievement in mathematics. *Procedia - Social and Behavioral Sciences*, 1(1), 953–957. <https://doi.org/10.1016/j.sbspro.2009.01.16>

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1994). *Self-efficacy*. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). Academic Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Betz, N. E., & Hackett, G. (1983). *Mathematics Self-Efficacy Scale* [Dataset]. In PsycTESTS Dataset. *American Psychological Association (APA)*. <https://doi.org/10.1037/t01563-000>

*Blake, M. E., Masten, W. G., Henley, T. B., & Ball, S. E. (2015). Sources and influence of mathematics self-efficacy in Jamaican college students. *Caribbean Journal of Psychology*, 7(1), 24-42.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2019). *Introduction to meta-analysis (2nd ed.)*. Wiley.

*Briley, J. S. (2012). The relationships among mathematics teaching efficacy, mathematics self-efficacy, and mathematical beliefs for elementary pre-service teachers. *Issues in the Undergraduate Mathematics Preparation of School Teachers: The Journal*, 5, 1–13.

Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement error of scores on the mathematics anxiety rating scale across studies. *Educational and Psychological Measurement*, 61(3), 373–386. <https://doi.org/10.1177/00131640121971266>

Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs type indicator score reliability across studies: A meta-analytic reliability generalization study. *Educational and Psychological Measurement*, 62(4), 590–602. <https://doi.org/10.1177/0013164402062004004>

Caruso, J. C., Witkiewitz, K., Belcourt-Dittloff, A., & Gottlieb, J. D. (2001). Reliability of scores from the Eysenck personality questionnaire: A reliability generalization study. *Educational and Psychological Measurement*, 61(4), 675–689. <https://doi.org/10.1177/00131640121971437>

*Cebollero, M. P., Fora, N. M., Permanyer, L. F., & Olmos, J. G. (2012). La evaluación de la ansiedad ante las matemáticas y su efecto como interferidor del rendimiento académico en la asignatura de estadística en psicología. *Revista del Congrés Internacional de Docència Universitària i Innovació (CIDUI)*, 1(1).

- Cousin, S. L., & Henson, R. K. (2000, January 27–29). *What is reliability generalization and why is important?* [Paper presentation]. Annual Meeting of the Southwest Educational Research Association, Dallas, TX.
- Crouch, M. (2016). *Reliability generalization: Exploring score reliability variance with Ryff's scale of psychological well-being* [Unpublished master's thesis]. Brock University.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- *Dobarro, R. V. (2007). *Solução de problemas e tipos de mente matemática: Relações com as atitudes e crenças de auto-eficácia* [Problem solving and types of mathematical mind: Relationships with attitudes and self-efficacy beliefs] [Unpublished doctoral dissertation]. Universidade Estadual de Campinas.
- Eurostat. (n.d.). *International Standard Classification of Education (ISCED)*. Retrieved March 8, 2024, from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International Standard Classification of Education \(ISCED\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED))
- *Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28(2), 161–186. [https://doi.org/10.1016/s0361-476x\(02\)00015-2](https://doi.org/10.1016/s0361-476x(02)00015-2)
- * Guàrdia-Olmos, J., Peró-Cebollero, M., Mancho-Fora, N., & Farràs-Permany, L. (2012). *Adaptació d'una bateria per a l'avaluació de prerequisits per a l'aprenentatge d'estadística en psicologia* [Adaptation of a battery for the assessment of prerequisites for learning statistics in psychology]. *REIRE: Revista d'Innovació i Recerca en Educació*, 5(1), 42–58.
- Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education*, 20(3), 261–273. <https://doi.org/10.5951/jresmetheduc.20.3.0261>
- Henson, R. K., & Hwang, D.-Y. (2002). Variability and prediction of measurement error in Kolb's learning style inventory scores: A reliability generalization study. *Educational and Psychological Measurement*, 62(4), 712–727. <https://doi.org/10.1177/0013164402062004011>
- Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies. *Measurement and Evaluation in Counseling and Development*, 35(2), 113–127. <https://doi.org/10.1080/07481756.2002.12069054>
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement*, 61(3), 404–420. <https://doi.org/10.1177/00131640121971284>
- Hess, T. J., McNab, A. L., & Basoglu, K. S. (2014). Reliability generalization of perceived ease of use, perceived usefulness, and behavioral intentions. *MIS Quarterly*, 38, 1-28. <https://doi.org/10.25300/MISQ/2014/38.1.01>
- Holland, D. F., Kraha, A., Zientek, L. R., Nimon, K., Fulmore, J. A., Johnson, U. Y., ... & Henson, R. K. (2018). Reliability generalization of the motivated strategies for learning questionnaire: a meta-analytic view of reliability estimates. *Sage Open*, 8(3). <https://doi.org/10.1177/2158244018802334>
- Kilgus, S. P., Eklund, K. R., Maggin, D. M., Taylor, C. N., & Allen, A. N. (2018). The student risk screening scale: A reliability and validity generalization meta-analysis. *Journal of Emotional and Behavioral Disorders*, 26(3), 143-155. <https://doi.org/10.1177/1063426617710207>
- Kranzler, J. H., & Pajares, F. (1997). An exploratory factor analysis of the Mathematics Self-Efficacy Scale revised (MSES-R). *Measurement and Evaluation in Counseling and Development*, 29(4), 215–228. <https://doi.org/10.1080/07481756.1997.12068906>
- Lane, G. G., White, A. E., & Henson, R. K. (2002). Expanding reliability generalization methods with KR-21 estimates: An RG study of the Coopersmith self-esteem inventory. *Educational and Psychological Measurement*, 62(4), 685–711. <https://doi.org/10.1177/0013164402062004010>
- Leach, L. F., Henson, R. K., Odom, L. R., & Cagle, L. S. (2006). A reliability generalization study of the self-description questionnaire. *Educational and Psychological Measurement*, 66(2), 285–304. <https://doi.org/10.1177/0013164405284030>
- Mahapoonyanont, N., Krahomwong, R., Kochakornjarupong, D., & Rachasong, W. (2010). Critical thinking abilities assessment tools: Reliability generalization. *Procedia - Social and Behavioral Sciences*, 2(2), 434–438. <https://doi.org/10.1016/j.sbspro.2010.03.038>

- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 575-596). Macmillan.
- Mji, A., & Alkhateeb, H. M. (2005). Combining reliability coefficients: Toward reliability generalization of the conceptions of mathematics questionnaire. *Psychological Reports*, 96(3), 627-634. <https://doi.org/10.2466/pr0.96.3.627-634>
- Moore, N. M. (2005). *Constructivism using group work and the impact on self-efficacy, intrinsic motivation, and group work skills on middle-school mathematics students* [Doctoral dissertation]. Capella University.
- Nilsson, J. E., Schmidt, C. K., & Meek, W. D. (2002). Reliability generalization: An examination of the career decision-making self-efficacy scale. *Educational and Psychological Measurement*, 62(4), 647-658. <https://doi.org/10.1177/0013164402062004007>
- Onwuegbuzie, A. J., & Daniel, L. G. (2000, November 17). *Reliability generalization: The importance of considering sample specificity, confident intervals, and subgroup differences*. [Paper presentation]. Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY, United States.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332. <https://doi.org/10.3102/00346543062003307>
- Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology*, 21(4), 325-344. <https://doi.org/10.1006/ceps.1996.0025>
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24(2), 124-139. <https://doi.org/10.1006/ceps.1998.0991>
- Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20(4), 426-443. <https://doi.org/10.1006/ceps.1995.1029>
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, 42(2), 190-198. <https://doi.org/10.1037/0022-0167.42.2.190>
- Ross, M. E., Blackburn, M., & Forbes, S. (2005). Reliability generalization of the patterns of adaptive learning survey goal orientation scales. *Educational and Psychological Measurement*, 65(3), 451-464. <https://doi.org/10.1177/0013164404272496>
- Schunk, D. H. (2009). *Learning theories: An educational perspective* (5th ed.). Pearson Education.
- Taylor, R. (2012). *Review of the motivated strategies for learning questionnaire (MSLQ) using reliability generalization techniques to assess scale reliability* [Doctoral dissertation]. Auburn University.
- Usher, E. L. (2009). Sources of middle school students' self-efficacy in mathematics: A qualitative investigation. *American Educational Research Journal*, 46(1), 275-314. <https://doi.org/10.3102/0002831208324517>
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20. <https://doi.org/10.1177/0013164498058001002>
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 44(3), 159-168. <https://doi.org/10.1177/0748175611409845>
- Vacha-Haase, T., Kogan, L. R., Tani, C. R., & Woodall, R. A. (2001). Reliability generalization: Exploring variation of reliability coefficients of MMPI clinical scales scores. *Educational and Psychological Measurement*, 61(1), 45-59. <https://doi.org/10.1177/00131640121971059>
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261-293. <https://doi.org/10.3102/10769986030003261>
- West African Examinations Council. (2005-2007). Chief examiners' report. Yaba, Lagos: WAEC.
- Williams, A. M., & Young, J. (2021). Reliability generalization meta-analyses in mathematics education research: A research synthesis. *International Journal of Education in Mathematics, Science and Technology*, 9(4), 741-759. <https://doi.org/10.46328/ijemst.1434>