Hawthorne effect and mathematical software based learning: a meta-analysis study

M Tamur1*, Y S Kusumah2*, D Juandi2, T T Wijaya3, A Nurjaman4, and A O Samura5

1Universitas Katolik Indonesia Santu Paulus Ruteng, Jl. Jend. A. Yani No. 10, Ruteng 86511, Indonesia
2Departemen Pendidikan Matematika, Universitas Pendidikan Indonesia, Jl. Dr. Sutomo No. 100, Bandung 40154, Indonesia
3Guangxi Normal University, China
4Institut Keguruan dan Ilmu Pendidikan Siliwangi, Cimahi, Indonesia
5Institut Agama Islam Negeri Ternate, Ternate, Indonesia

*maximustamur@unikastpaulus.ac.id, yayaskusumah@yahoo.com

Abstract. The use of mathematical software in learning has been widely applied. This meta-analysis study was conducted to examine the "Hawthorne effect", or the novelty of using mathematical software in learning. Related databases are sought for studies that qualify for examining the effects of using mathematical software compared to traditional teaching on students' mathematical abilities. Random-effect models with 95% confidence intervals (CIs) were estimated. 40 effect sizes were analyzed from 40 primary studies that have been published in journals, or national and international proceedings from 2010 to 2020, and a total of 3072 students were included in this meta-analysis. The Comprehensive Meta-Analysis (CMA) program is used to assist analysis. As a result of the study, the overall effect size was 0.886 with a standard error of 0.092. The results of the follow-up analysis showed that the duration of the treatment was related to the use of mathematical software. It was found that the use of mathematical software given in the span of 0 - 4 weeks (effect size = 1.193) was more effective if carried out in more than 4 weeks (effect size = 0.535). Thus the Hawthorne effect influences the effectiveness of the use of mathematical software on students' mathematical abilities. This fact is considered by education practitioners in using mathematical software in the future.

1. Introduction
The use of mathematical software in teaching is increasingly gaining attention from teachers and researchers [1,2]. The use of mathematical software is becoming a trend and has triggered a flurry of study widely [3-7]. The use of mathematical software is considered to affect students' mathematical abilities. However, previous studies investigating these theoretical assumptions have shown mixed results. Research conducted by [6-9] has shown that the use of mathematical software is effective in improving students' mathematical abilities rather than conventional learning. However, [10,11] has found that the mathematical abilities of students who are taught using mathematical software are no better than the abilities of students who are taught using conventional learning. On the other hand,
there is a need that education practitioners need accurate information to decide under what conditions the use of mathematical software will be effective in improving students’ mathematical abilities.

Highlighting the variations in the results of these studies has led some researchers to conduct meta-analysis research [12-16]. The research that has been carried out focuses on evaluating the use of computer and mathematical software in learning by examining variations in results from primary studies. It has been found that mediator variables such as study years, education levels, sample sizes, and duration of treatment, are related to the use of mathematical software in learning. To date, there have been no studies that specifically examined the Hawthorne effect related to the duration of treatment. The Hawthorne effect, or novelty, occurs when students are encouraged to make greater efforts simply because of the novelty of treatment. If treatment is continued for a long time, it will lose its effectiveness over time and the results will become less visible [16].

In our previous research, we have examined the Hawthorne effect of applying realistic mathematics learning to students’ mathematical abilities. It was found that Hawthorne has affected students’ mathematical abilities [17]. The purpose of this study is to find the overall effect of learning supported by the use of mathematical software and examine the effects of Hawthorne so as to provide conclusive conclusions to practitioners of Education in implementing the use of mathematical software in the future.

2. Methods

2.1 Research design
The design of this study is a meta-analysis that combines two or more primary studies that have been published to integrate the findings [18,19]. The stages in the meta-analysis are: First, inclusion criteria for studies included in the meta-analysis will be explained. Second, the procedure for collecting empirical data and coding of study variables will be explained. Third, statistical techniques to investigate the relationship between study variables and effect sizes [20,21]. This research follows these stages.

2.2 Inclusion Criteria
Inclusion criteria namely the eligibility of standards used in selecting empirical data namely; (a) in the form of a Journal, Proceedings, Thesis, and dissertation; (b) is the result of research in Indonesia concerning the use of mathematical software on students' mathematical abilities; (c) publication in at least the last decade (2010-2020); and (e) has statistical information for the effect size transformation namely; average, standard deviation and sample size.

2.3 Data collection
Empirical data in this study were obtained from published articles and abstracts identified by a computer literature search from an electronic database that included ERIC document, SAGE Publishing and Google Scholar. Empirical data search was also obtained by searching manuals from library sources for theses and dissertations for one year. Cross-reference checks have been carried out through email communication with primary study researchers. This stage found 134 studies examining the effect of using mathematical software. Based on the inclusion criteria, 40 primary studies that met the eligibility to be analyzed. The instrument used in this study was the coding sheet. The coding process in addition to assisting researchers in analyzing data also avoids forgotten data. Coding in data analysis includes information extracted from primary studies, namely the duration of statistical information and the duration of treatment to find the Hawthorne effect. To ensure data is entered without error, the two encoders have filled out the coding form separately and then compared.

2.4 Statistic analysis
The statistical analysis in this study is based on suggestions [20] namely (a) calculating the effect size for each primary study and the combined effect size; (b) conducting heterogeneity tests and selecting
estimation models; (c) check publication bias; and (d) calculate the p-value to test the research hypothesis. The examination of variations in effect size is to analyze the treatment duration variable, once the model estimated is randomized. The application that helps with data analysis is the Comprehensive Meta-Analysis (CMA) program. The Hedge's g equation is used to determine the effect size index. Interpretation of effect sizes, using the Thalheimer & Cook classification as shown in Table 1 below:

| Range of Effect Size (ES) | Interpretation       |
|--------------------------|----------------------|
| -0.15 ≤ ES < 0.15        | no effect            |
| 0.15 ≤ ES < 0.40         | Low effect           |
| 0.40 ≤ ES < 0.75         | Medium effect        |
| 0.75 ≤ ES < 1.10         | High effect          |
| 1.10 ≤ ES < 1.45         | Very high effect     |
| ES ≥ 1.45                | excellent effect     |

The CMA program can calculate the Z value used to test the significance of the influence of mathematical software on students' mathematical abilities, and provide an average effect size with confidence intervals, as well as homogeneity between groups namely the Qb value. If $Z_{\text{count}} > Z_{\text{table}}$ with $p < 0.05$ then the null hypothesis is rejected [20]. This means that the use of mathematical software produces a positive effect size on students' mathematical abilities compared to conventional approaches. The most important decision that must be made when conducting a meta-analysis is whether to use a fixed-effect or random-effect estimation model [22].

The fixed-effect model signifies that the size of the effect between studies or study groups is homogeneous. The random-effects model is used when the effect size is statistically heterogeneous ($Qb > \chi^2_{2.95}; p < 0.05$) then the hypothesis stating the effect size between studies is homogeneous is rejected. Rejecting Qb means that the effect size between studies or study groups may not measure the same population parameters [20]. In other words, there is a statistically significant difference in the average effect size for each group of study characteristics [16].

3. Result and Discussion

The first objective of this meta-analysis research is to find a measure of the overall effect of learning using mathematical software on students' mathematical abilities. Based on the overall calculation, Figure 1 presents the spread and interpretation of effect sizes based on classification [23].

![Figure 1. Interpretation of Effect Size](image)

Figure 1 shows that the distribution of research was almost evenly distributed into 6 groups. This means that the effects of using mathematical software are not consistent in any of these classifications. This finding shows the variation of results in mathematical software on students’ mathematical abilities.
The next step is to test the heterogeneity of effect sizes between studies using Q statistics. Table 2 shows a comparison of meta-analysis results in accordance with the estimation model.

Table 2. Comparison of Meta-Analysis Results According to Estimates

| Model Estimate      | n  | Qb   | df (Q) | Nilai-p |
|---------------------|----|------|--------|---------|
| Model Efek-Tetap    | 40 | 202.232 | 39     | 0.000   |
| Model Efek-Aacak    | 40 |        |        |         |

Table 2 shows that the Qb value of 202.232. This value is found to be greater than 54.572 in degrees of freedom 39 in table $\chi^2$. Thus, the distribution of effect sizes was found to be heterogeneous. Therefore the estimation model used is a random-effects model. Next, a publication bias analysis is performed based on a random-effects model, which is checking whether there is a tendency for journals to publish only significant studies that lead to meta-analyses that are too high for the actual effect size. [20]. Publication bias can be determined by examining the funnel plot. There is no bias if the spread of effect sizes from primary studies shows symmetrical distribution around symmetric lines in the funnel plot [20]. However, just looking at a funnel plot is not enough, because everyone has a different perspective. One approach that helps to clarify the presence or absence of publication bias is to use Rosenthal's fail-safe N (FSN) statistics [24]. If the value of FSN / (5k + 10)> 1 where k is the number of studies included in the meta-analysis then this research is resistant to publication bias [25]. Figure 2 shows a research funnel plot.

Figure 2. Research Funnel Plots

Figure 2 shows that the spread of the effect size is not completely symmetrical around the vertical line. Therefore Rosenthal's fail-safe N (FSN) statistics are evaluated. The following table 3 shows the N tests.

Table 3. Rosenthal's fail-safe N (FSN) statistics

| Bias condition                  |     |
|--------------------------------|-----|
| Z value for the observed study  | 21.45163 |
| P value for the observed study  | 0.00000 |
| Alpha                          | 0.05 |
| Tails                          | 2    |
| Z value for Alpha              | 1.95996 |
| Number of Observed Studies     | 40   |
| FSN                            | 4752 |

Table 3 shows that the N value calculated was 4752. Based on the formula, the calculation result of $4752 / (5 * 40 + 10)$, was 19.008> 1. Based on this calculation the study included in this analysis was
resistant to publication bias. That is, no studies were lost or needed to be added to the analysis as a result of publication bias.

The next step is to calculate the p-value to test the research hypothesis, which is to evaluate the effect of using mathematical software as a whole compared to conventional approaches. Table 4 below shows the results of the overall analysis.

Table 4. Overall Analysis Results Based on the Random-Effect Model

| Model      | n  | Z       | p      | Effect Size | Standard Error | 95% Confidence Interval |
|------------|----|---------|--------|-------------|-------------------|------------------------|
| Efek-Aacak | 40 | 9.582   | 0.0000 | 0.886       | 0.092             | 0.705 - 1.067          |

Based on Table 4, the overall effect size is 0.886 with the lower limit of the 95% confidence interval is 0.705 and the upper limit is 1.067. This effect size is accepted as a high effect. The results of the z test calculations to determine statistical significance, z scores were found 9.582 with p = 0.000. This result can be said to be statistically significant at the level of p <0.001. The size of the effect shows that the average student exposed to mathematical software exceeds the mathematical ability of 76% of students in conventional classes that are initially equivalent.

This finding is consistent with the results of previous meta-analysis studies as conducted by [12] and [13], with combined effect sizes of 1.02 and 0.84, respectively. The researchers conducted a meta-analysis to examine the effectiveness of using mathematical software on student academic achievement. Although the number of studies analyzed in this study is different from previous studies, the findings show an almost similar overall trend. However, other meta-analyses show slightly different results as reported by [14-16], where the combined effect sizes are 0.449, 0.33, and 0.273, respectively. This difference in results is quite large so it needs to be further investigated by involving more studies analyzed to check its consistency.

Finally, a follow-up analysis to examine the Hawthorne effect is performed after it is known that the random-effects model as the estimated model is chosen. This means that the variation in effect size is evaluated from the identified mediator variable, which is the duration of the treatment. Table 5 shows a summary of the results of the analysis.

Table 5. Summary of Analysis Results Based on Primary Study Mediator Variables

| Duration of Treatment | Effect Size | Test of null | Test of Homogeneity |
|-----------------------|-------------|--------------|---------------------|
|                       | Mean | SE  | Z     | p     | Q-value | df(Q) | p     |
| 0-4 weeks             | 1.193| 0.066| 18.214| 0.000| 49.943  | 15    | 0.000 |
| More than 4 weeks     | 0.535| 0.057| 9.461 | 0.000| 97.627  | 16    | 0.000 |
| Unspecified           | 0.605| 0.080| 7.578 | 0.000| 8.590   | 6     | 0.198 |
| Within Classes        |       |      |       |      | 201.785 | 37    | 0.000 |
| Between Classes       |       |      |       |      | 7.358   | 2     | 0.000 |

Table 5 shows that the Z scores for all study groups were found to be greater than Ztable at the level of p <0.001. This indicates the advantages of learning using mathematical software. The Qb statistical value obtained as a result of the homogeneity test was 7.358 greater than the value of 5.991 at a 95% confidence level, a significance level of 0.05, and a degree of freedom 2. That is, there was a significant difference in effect sizes between groups according to the duration of treatment. The combined effect size of the study group with a treatment duration of 0-4 weeks was found to be greater than the study group with a treatment duration of more than 4 weeks. This means that there is evidence that the effect size of the study group with a shorter treatment duration is better than the study group with a longer treatment duration.

This finding is surprising because it contradicts the notion that mathematical software increases retention and attracts students’ interest in learning. This finding clearly shows that the Hawthorne
effect also occurs in learning that is supported by the use of mathematical software. It is clear that students will feel happy with the new method but will get bored with time. As a result, the effects of using mathematical software become less clear when the duration of the treatment is longer. Similar findings were reported by [16] and [17] who each conducted a meta-analysis of the effectiveness of computer-based learning and realistic mathematics learning. However, this finding is tentative because there are still study groups that do not report the duration of the treatment.

4. Conclusion
This meta-analysis study was conducted to examine the effect of the use of mathematical software on students' mathematical abilities and examine the effects of Hawthorne that occur. Based on a random-effects model with a 95% confidence interval, the combined effect size was found to be 0.886 with a standard error of 0.92. These findings indicate the use of mathematical software produces a greater effect size of students' mathematical understanding abilities than conventional approaches. An examination of the Hawthorne effect is done by evaluating the relationship between the duration of the treatment and the effectiveness of using mathematical software. It was found that the use of mathematical software was more effective in the condition of the duration of treatment from 0–4 weeks. There is evidence that continued treatment for a long time will lose its effectiveness over time and the results will be less visible. This study shows the use of mathematical software has a high positive effect on students' mathematical abilities. However, this finding is only supported by primary studies containing statistical information for transforming effect sizes.

5. References
[1] Koparan T, 2015 The effect of dynamic geometry software on prospective teachers’ achievement about locus problems Turkish Online J. Educ. Technol. 2 p. 587–591.
[2] Adelabu F M Makgato M and Ramaligela M S, 2019 The Importance of Dynamic Geometry Computer Software on Learners ‘ Performance in Geometry Electron. J. e-Learning 17, 1 p. 52–63.
[3] Baki A Kosa T and Guven B, 2011 A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualisation skills of pre-service mathematics teachers Br. J. Educ. Technol. 42, 2 p. 291–310.
[4] Adelabu F M and Makgato M, 2019 Attitudes of male and female students to dynamic geometry computer software for learning mathematics World Trans. Eng. Technol. Educ. 17, 3 p. 314–319.
[5] Gordon S P and Gordon F S, 2020 Visualizing and Understanding Hypothesis Testing Using Dynamic Software Primus 30, 2 p. 172–190.
[6] Juandi D and Priatna N, 2018 Discovery learning model with geogebra assisted for improvement mathematical visual thinking ability J. Phys. Conf. Ser. 1013, 1.
[7] Kusumah Y S Kustiawati D and Herman T, 2020 The Effect of GeoGebra in Three-Dimensional Geometry Learning on Students’ Mathematical Communication Ability Int. J. Instr. 13, 2 p. 895–908.
[8] Hartatiana H Darhim D and Nuraelah E, 2017 Improving Junior High School Students’ Spatial Reasoning Ability Through Model Eliciting Activities with Cabri 3D Int. Educ. Stud. 11, 1 p. 148.
[9] Subroto T, 2011 The use of Cabri 3D Software As Virtual Manipulation Tool in 3-Dimension Geometry Learning To Improve Junior High School Students’ Spatial Ability Int. Semin. Fourth Natl. Conf. Math. Educ. July p. 609–618.
[10] Nuraeni Z and Rosyid A, 2019 Pengaruh Model Pembelajaran Index Card Match (ICM) dengan Problem Posing Berbantuan Software MATLAB terhadap Kemampuan Pemecahan Masalah J. Elem. 5, 1 p. 12.
[11] Priyono S and Hermanto R, 2015 Peningkatan kemampuan representasi matematik peserta didik dengan menggunakan model problem based learning (PBL) berbantuan media software
Geogebra 1, 1 p. 55–64.

[12] Chan K K and Leung S W, 2014 Dynamic geometry software improves mathematical achievement: Systematic review and meta-analysis J. Educ. Comput. Res. 51, 3 p. 311–325.

[13] Demir S and Başol G, 2014 Effectiveness of computer-assisted mathematics education (CAME) over academic achievement: A meta-analysis study Kuram ve Uygulamada Egit. Bilim. 14, 5 p. 2026–2035.

[14] Clark D B Tanner-Smith E E and Killingsworth S S, 2016 Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis Rev. Educ. Res. 86, 1 p. 79–122.

[15] Anıl Ö Batdi V and Küçükozer H, 2018 The effect of computer-supported education on student attitudes: A meta-analytical comparison for the period 2005-2015 Kuram ve Uygulamada Egit. Bilim. 18, 1 p. 5–22.

[16] Bayraktar S, 2001 A meta-analysis of the effectiveness of computer-assisted instruction in science education J. Res. Technol. Educ. 34, 2 p. 173–188.

[17] Tamur M Juandi D and Adem A M G, 2020 Realistic Mathematics Education in Indonesia and Recommendations for Future Implementation: A Meta-Analysis Study J. Teor. dan Apl. Mat. 4, 1 p. 17–27.

[18] Glass G V., 1976 Primary, Secondary, and Meta-Analysis of Research Educ. Res. 5, 10 p. 3–8.

[19] Hunter J E and Schmidt F L, 2004 Methods of meta analysis: Correcting error and bias in research finding.

[20] Borenstein M Hedges L V and Rothstein H R, 2009 Introduction to Meta-Analysis. Chicago, IL, USA: Springer is part of Springer Science+Business Media.

[21] Pigott T D, 2012 Advances in Meta-Analysis. Chicago, IL, USA: Springer is part of Springer Science+Business Media.

[22] Haidich A B, 2010 Meta-analysis in Medical Research Hippokratia 14, Supp 1 p. 29–37.

[23] Thalheimer W and Cook S, 2002 Effect Sizes.pdf5.pdf Work Learn. Research August p. 1–9.

[24] Turgut S and Turgut I G, 2018 The Effects of Cooperative Learning on Mathematics Achievement in Int. J. Instr. 11, 3 p. 663–680.

[25] Mullen B Muellerleile P and Bryant B, 2001 Cumulative meta-analysis: A consideration of indicators of sufficiency and stability Personal. Soc. Psychol. Bull. 27, 11 p. 1450–1462.

Acknowledgments
This research was supported by the Santu Paulus Ruteng Foundation