Sustained Magma Use and Year 6 Mathematics Results in Sweden

A school-level analysis of national test results, 2018/19–2024/25

Martina Galfrè
Data Analyst, Magma Math

October 2026

Abstract

This paper examines how the use of Magma relates to results on the Year 6 national exam in mathematics in Sweden. Skolverket's published school-level results for the five national exams held between 2018/2019 and 2024/2025 were linked to Magma usage data at school level. The main finding concerns sustained use. Comparing each school with its own results over time, across 2,791 schools with results in at least three of the five available exam years, including 2024/2025, years with sustained use of Magma (40 or more active weeks) were associated with 0.21 higher average grade points ($p<0.01$) and approximately one fewer failing student per 100 ($p = 0.03$), after accounting for stable differences between schools and county-specific changes over time. A separate analysis counting only activity before May also found a positive and statistically significant association. Two further comparisons point in the same direction. Schools with high use in their first year gained 0.30 grade points more than schools that never used Magma ($p<0.01$), with no statistically detectable difference in the available pre-adoption comparison. Over the full period, schools that started using Magma in 2021/2022 and continued to use it moved from a Fail Rate above that of schools that never used Magma to one below it; their Fail Rate improved by 3.0 percentage points more than that of schools that never used Magma. Together, the results provide encouraging evidence that sustained use of Magma is linked to stronger mathematics results.

1 Introduction

Midway through compulsory school, all Swedish students take the national exam in mathematics in Year 6, which assesses their knowledge against a common national standard. Among students with a recorded Year 6 mathematics national-test grade in 2024/2025, 13.6% received F[1]. The share of students graded F is referred to below as the Fail Rate.

Magma is a digital mathematics platform used in Swedish schools for assignments and formative assessment. Because Skolverket publishes exam results for every school, and Magma records usage for every school that uses it, the analysis can draw on nationwide school-level data rather than a selected sample of schools. This paper addresses three questions: whether results vary with how regularly Magma is used; whether differences are visible already in the first year of use; and how the Fail Rate of schools using Magma developed over time compared with schools that did not use Magma.

2 Methodology

2.1 Data collection

Skolverket publishes the results of the Year 6 national exam in mathematics for each school[2], including the number of students with an exam grade, the share of students achieving a passing grade (A–E) and the average result in grade points (E = 10, D = 12.5, C = 15, B = 17.5, A = 20, F = 0). Data is available for 2018/2019 and 2021/2022 to 2024/2025; no national exams were held in 2019/2020 and 2020/2021.

Each school's use of Magma was measured as the number of assignment problems solved by students each week, and the number of students active in Magma. Usage data was matched to Skolverket's results using the school unit code.

Skolverket reports the pass rate of schools with 1–4 failing students as "~100". The comparisons of Fail Rates between groups of schools (the Fail Rate over time and the comparison of the 2022/2023 starters) treat these as 100% passing for calculation. The within-school Fail Rate results count each such school as having 2.5 failing students, the midpoint of the range, so that schools with a few failing students are not recorded as having none. Results under the alternative treatment are reported in Appendix Tables A6 and A8. Grade-point averages are not affected by this rule. Skolverket does not publish results based on very few students; these school-years are not included, for schools using Magma and schools that never used Magma alike.

[1] Skolverket (2025). Resultat på de nationella proven i årskurs 3, 6 och 9, läsåret 2024/25, Diagram 13.

[2] Skolverket. Grundskola – Resultat nationella prov årskurs 6, skolnivå, matematik, läsåren 2018/19 och 2021/22–2024/25.

2.2 Definitions

Table 1. Definitions used in the analysis

Term	Definition
Fail Rate	Share of students graded F on the exam
Active week	A week in which students at the school solved at least 20 assignment problems
Magma school	A school with at least 20 active weeks and 50 problems per student every year from the first year of use, and fewer than 10 active weeks in the year before
Problems per student	Assignment problems solved during the school year divided by the number of students active in Magma at the school
High use	At least 30 active weeks and 100 problems per student in a school year
Weekly use	Fewer than 30 active weeks; 30–39 active weeks; 40 or more active weeks (sustained use). Active weeks are counted per calendar week across the school year
Never used Magma	Fewer than 5 active weeks in every school year from 2017/2018 to 2024/2025

2.3 Analysis

Schools decide how much to use Magma, and differences in results between users and non-users could therefore reflect differences between schools. To account for this, we used three comparisons, each controlling for a different source of difference:

- **Each school compared with itself.** A regression with school and year fixed effects compares results in years with Magma with the same school's results in years without Magma. It includes 2,791 schools with results in at least three of the five available exam years, including 2024/2025, municipal and independent. Schools whose usage does not vary over the period contribute to the estimation of the year effects. Stable differences between schools, such as student population, area and resources, do not affect this comparison. The main specification compares schools only with schools in the same county in the same year, accounting for stable school differences and county-specific changes over time.
- **The first year of use.** For each municipal school starting Magma between 2021/2022 and 2023/2024, the change from the last exam before Magma to the first exam with Magma is compared with the change for schools that never used Magma over the same two exams, controlling for each school's previous result.
- **The Fail Rate over time.** Municipal schools that started using Magma in 2021/2022 are compared with municipal schools that never used Magma, at the last exam before Magma and at the latest exam. The comparison assumes that national changes in results would otherwise have affected both groups equally, so that the difference between their changes reflects what differs between the groups.

Unless stated otherwise, averages and regressions are weighted by the number of students with an exam grade, and standard errors are clustered by school. Two robustness checks differ: the matched-pair comparison averages the differences between school pairs without weights and is tested with a t-test, and the permutation test derives its p-value from random reassignments. Descriptive comparisons, such as pass rates and Fail Rates by year, are reported as published; differences between groups are reported with p-values or 95% confidence intervals. The study is correlational, with statistical controls for differences between schools.

Additional checks examined sensitivity to comparison methods, usage definitions and timing, as applicable: pairing each school with a similar school that never used Magma, comparing

schools only within the same county, alternative definitions of high use, counting only activity before May, a permutation test, and a comparison of the years before schools started using Magma. The results of these checks are reported in the Results section and the Appendix.

3 Results

3.1 Sustained use

Across the 2,791 schools, results were higher in years with more active weeks of Magma than in the same school's years without Magma. In years with fewer than 30 active weeks, no statistically detectable difference was found (+0.03, $p = 0.59$). In years with 30–39 active weeks, results were 0.11 grade points higher ($p = 0.06$), and in years with sustained use of 40 or more active weeks, 0.24 grade points higher ($p<0.01$; Figure 1). Taken together, years with at least 30 active weeks had results 0.16 grade points higher ($p<0.01$). When each school is also compared only with schools in the same county in the same year, accounting for county-specific changes over time, sustained use was associated with 0.21 higher grade points ($p<0.01$; Table 2).

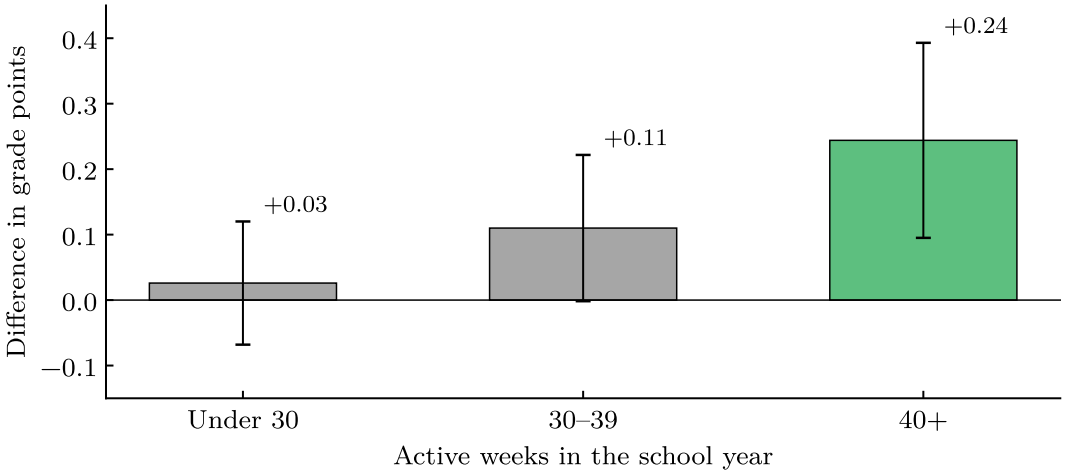


Figure 1. Results by weekly use, compared with the same school's years without Magma. The figure shows the difference in average grade points between years with Magma and the same school's years without Magma, by number of active weeks in the school year, from a regression with school and year fixed effects. Based on 4,174 school-years with Magma and 8,730 school-years without. Error bars show 95% confidence intervals.

The same pattern is seen in the Fail Rate. In years with sustained use, 1.2 fewer students per 100 failed than in the same school's years without Magma ($p<0.01$). For the 513 schools with at least one such year, this corresponds to a Fail Rate of 11.6% instead of 12.8%, or 9% fewer failing students (Figure 2). These Fail Rate results count "~100" as 2.5 failing students.

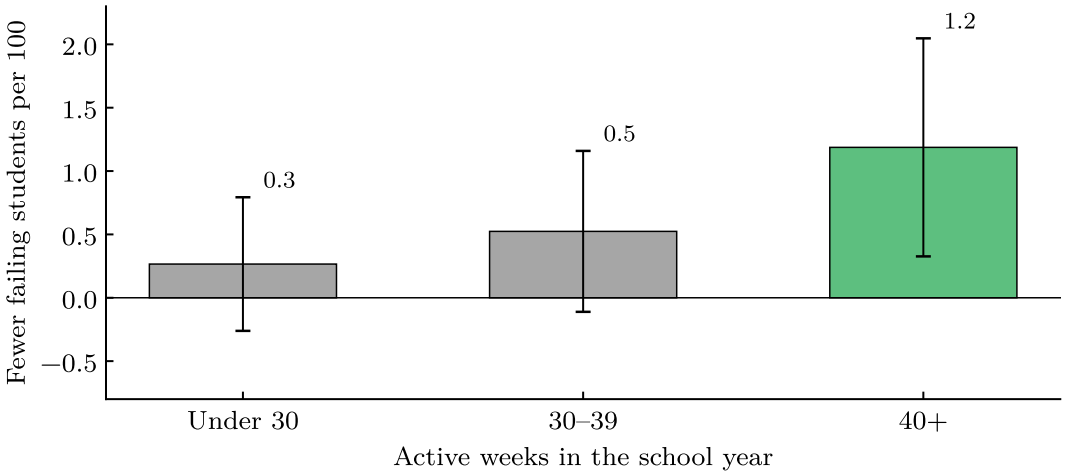


Figure 2. Fail Rate by weekly use, compared with the same school's years without Magma. The figure shows how many fewer students per 100 failed in years with Magma than in the same school's years without Magma, by number of active weeks in the school year, from a regression with school and year fixed effects, counting "~100" as 2.5 failing students. Error bars show 95% confidence intervals.

Table 2 reports both specifications. With county-by-year controls, the pattern is unchanged: years with sustained use have +0.21 grade points ($p<0.01$) and 1.0 fewer failing students per 100 ($p = 0.03$). Counting "~100" as no failing students gives a similar reduction, 1.0 fewer per 100, which is not statistically significant at the 5% level with county-by-year controls ($p = 0.07$; Appendix Table A8). The comparison is based on 8,730 school-years without Magma.

Table 2. Results in years with Magma compared with the same school's years without Magma

Active weeks	School-years	Grade points	Grade points, county × year	Failing per 100	Failing per 100, county × year
Under 30 active weeks	1,822	+0.03	+0.03	−0.27	−0.25
30–39 active weeks	1,477	+0.11	+0.10	−0.52	−0.43
40+ active weeks	875	+0.24**	+0.21**	−1.19**	−1.01*

School and year fixed effects; county × year: school and county-by-year fixed effects. Failing per 100 counts "~100" as 2.5 failing students. Standard errors clustered by school. ** $p<0.01$, * $p<0.05$.

As a separate sensitivity analysis, activity was counted only in weeks before May. Years with at least 30 active weeks before May were associated with 0.17 higher grade points ($p<0.01$) and 0.8 fewer failing students per 100 ($p<0.01$) than the same school's years without Magma, and with 0.15 higher grade points ($p = 0.01$) when schools were compared only within the same county and year. The threshold differs from the full-year definitions above, so the estimates are not directly comparable, but the association remains positive and statistically significant.

3.2 The first year of use

Among the schools that started using Magma between 2021/2022 and 2023/2024, the 308 schools with high use in their first year achieved 0.30 grade points more than schools that never used Magma, comparing the last exam before Magma with the first exam with Magma ($p<0.01$). The 161 schools with lighter use in their first year showed no statistically detectable difference compared with schools that never used Magma (+0.09, $p = 0.51$; Figure 3).

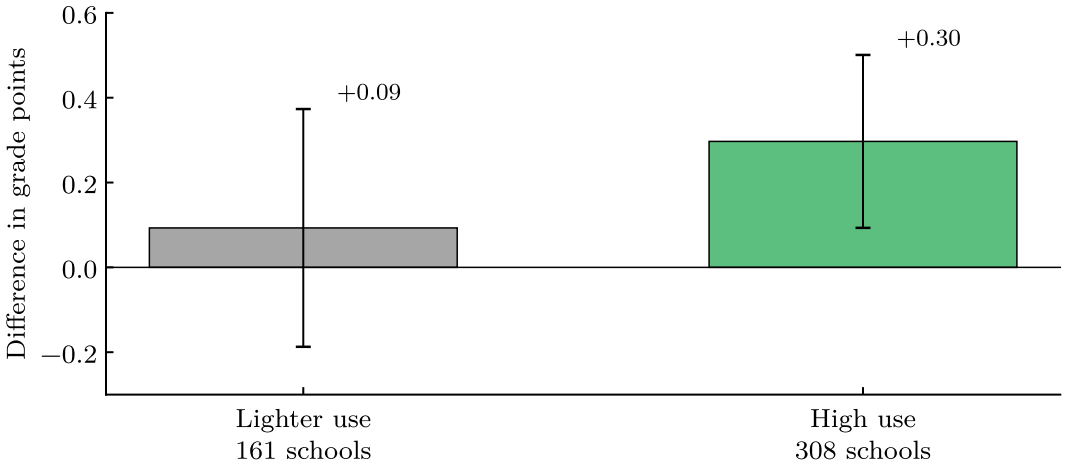


Figure 3. Difference compared with schools that never used Magma, first year of use. The figure shows the change in average grade points from the last exam before Magma to the first exam with Magma, minus the change for schools that never used Magma over the same two exams. High use: at least 30 active weeks and 100 problems per student in the first year. Error bars show 95% confidence intervals.

Schools that chose to start Magma could already have been improving before they started. To test this, we compared the change in results of the same schools in the years before they started Magma (2018/2019 to 2021/2022) with that of schools that never used Magma. Before starting, there was no statistically detectable difference (0.00 grade points, $p = 0.98$); in the first year of use, the difference was +0.30 grade points ($p<0.01$; Figure 4).

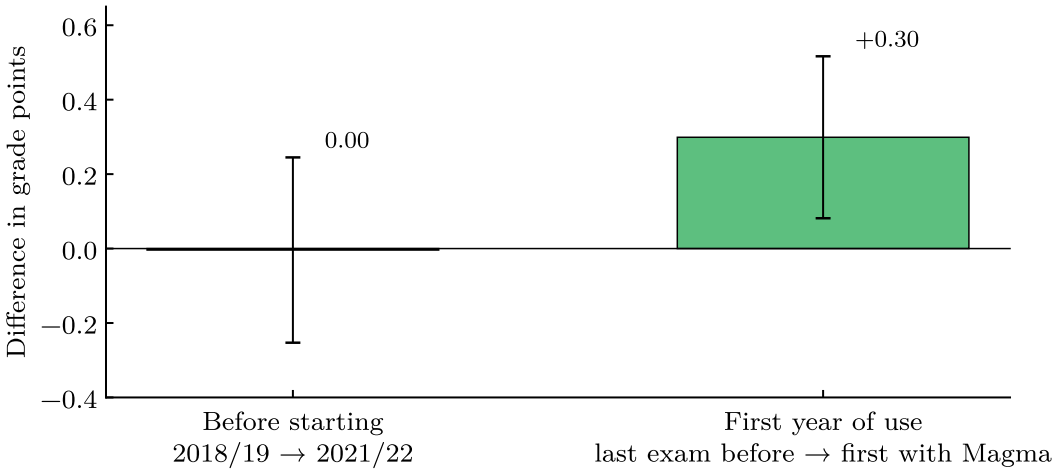


Figure 4. Change compared with schools that never used Magma, before and after starting. The figure shows, for schools that started using Magma in 2022/2023 or 2023/2024 with high use in the first year, the change in average grade points relative to schools that never used Magma: in the years before they started ($p = 0.98$) and in their first year of use ($p<0.01$). Both comparisons control for each school's previous result. Error bars show 95% confidence intervals.

The first-year result was tested with several alternative methods (Figure 5 and Appendix Table A3). In each case the difference is positive. When schools are compared only with schools in the same county, the estimate is +0.18 grade points ($p = 0.09$). In a permutation test, the high-use label was reassigned at random among the schools 500 times; a difference as large as the observed one occurred in 1 of 500 cases ($p = 0.004$). The definition of a Magma school

requires continued use after the first year; without this requirement, the first-year difference for the 328 schools with high use is +0.28 grade points ($p<0.01$).

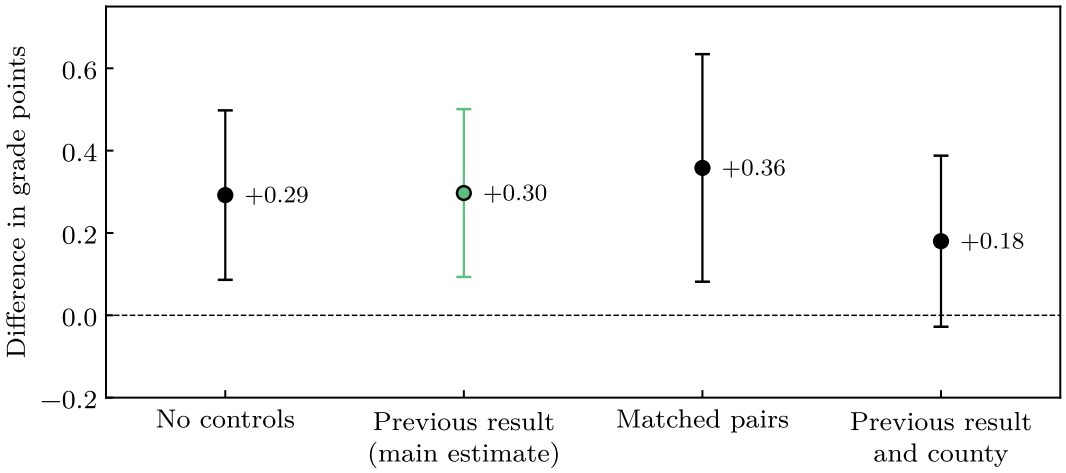


Figure 5. First-year difference by method of comparison, grade points. The figure shows the first-year difference between schools with high use and schools that never used Magma, estimated with four methods: without controls; controlling for each school's previous result (main estimate); pairing each school with the school that never used Magma closest in previous result and size, in the same years; and controlling for both previous result and county. Error bars show 95% confidence intervals.

The result does not depend on the exact definition of high use. Estimates are between 0.23 and 0.31 grade points for all definitions from 20 to 30 active weeks and from 50 to 200 problems per student (all $p<0.01$; Figure 6 and Appendix Table A4).

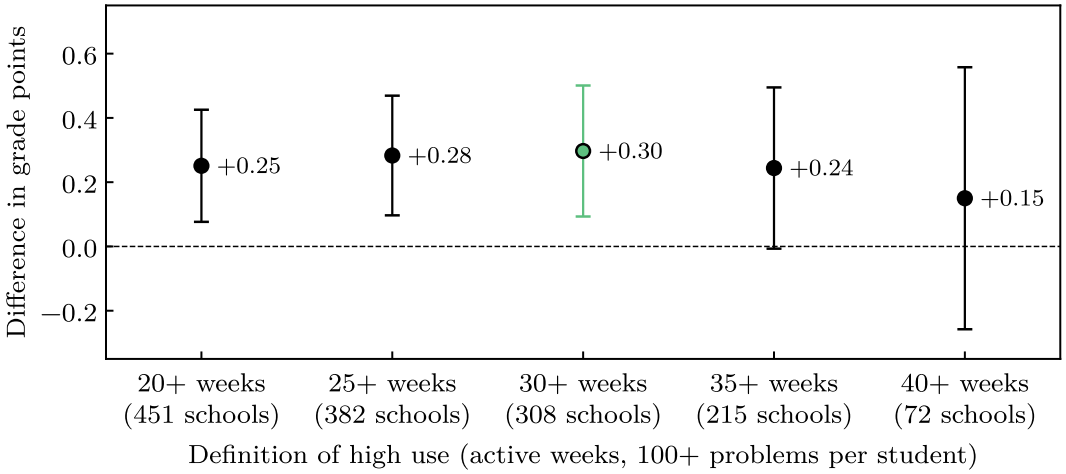


Figure 6. First-year difference by definition of high use, grade points. The figure shows the first-year difference compared with schools that never used Magma when high use is defined as at least 20, 25, 30, 35 or 40 active weeks with at least 100 problems per student. The main analysis uses 30 weeks. Fewer schools reach the stricter definitions, which widens the confidence intervals. Error bars show 95% confidence intervals.

Estimates vary between start years (Appendix Table A2): +0.32, +0.38 and +0.21 grade points for schools that started in 2021/2022, 2022/2023 and 2023/2024 respectively. For schools that started in 2024/2025, whose continued use cannot yet be observed, the estimate was −0.17

and not statistically significant; including them gives a pooled estimate of +0.15 ($p = 0.10$; Appendix Table A3).

3.3 The Fail Rate over time

In 2018/2019, before Magma, the schools that started using Magma in 2021/2022 had a Fail Rate of 12.7%, compared with 11.7% for schools that never used Magma. In 2024/2025 the Fail Rate of the Magma schools had decreased to 10.3%, while it had increased to 12.4% for schools that never used Magma. The difference between the two changes was 3.0 percentage points, equivalent to approximately 24% of the Magma schools' baseline Fail Rate (Figures 7 and 8). Alternative calculations are reported in Appendix Table A6.

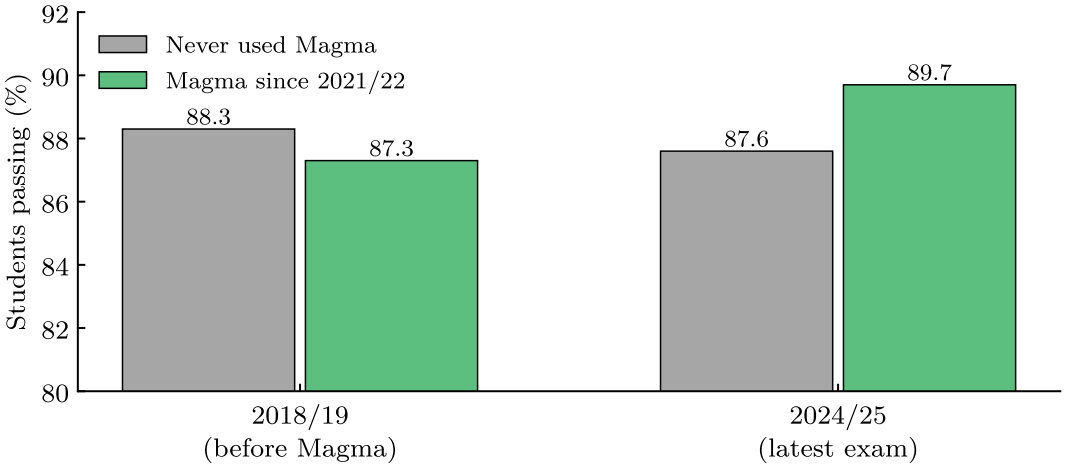


Figure 7. Pass rate on the Year 6 national exam in mathematics. The figure shows the share of students achieving a passing grade (A–E) in schools that started using Magma in 2021/2022 and continued to use it (89 schools), and in schools that never used Magma (1,137 schools), before Magma (2018/2019) and at the latest exam (2024/2025), based on the results published by Skolverket. Note that the vertical axis starts at 80%.

The Fail Rate at each of the five exams shows how the difference developed (Figure 8). Before Magma, the Fail Rate of the Magma schools was 1.0 percentage points higher than that of schools that never used Magma. The difference closed in the first year of use and widened in the following years, to 0.7 percentage points lower in 2023/2024 and 2.1 percentage points lower in 2024/2025.

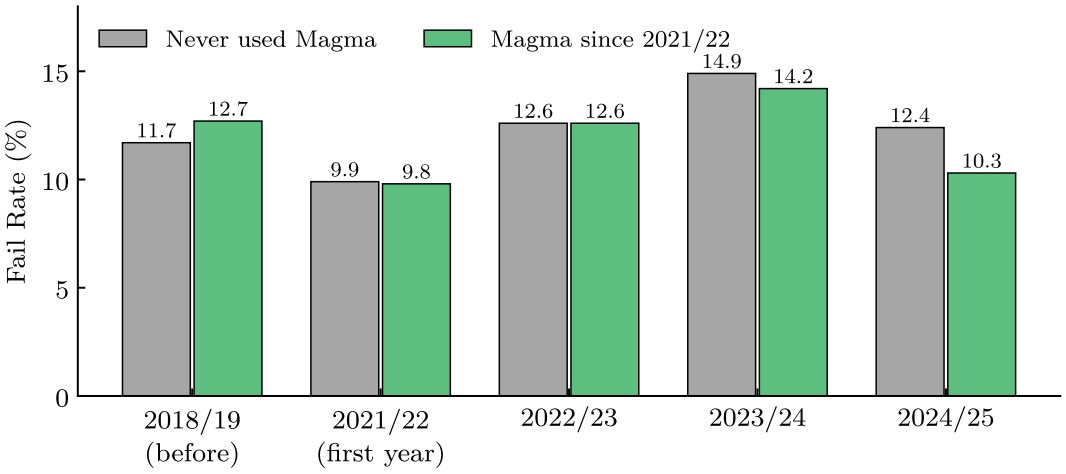


Figure 8. Fail Rate at each national exam. The figure shows the Fail Rate at each of the five national exams for schools that started using Magma in 2021/2022 and continued to use it, and for schools that never used Magma, based on the results published by Skolverket. No national exams were held in 2019/2020 and 2020/2021. The difference between the groups (Magma minus never-users) was +1.0, −0.1, 0.0, −0.7, −2.1 percentage points respectively.

A shorter comparison of the same 190 schools that started Magma in 2022/2023, before and after their first year, shows a Fail Rate 1.9 percentage points lower than for schools that never used Magma, equivalent to 20% fewer failing students ($p = 0.04$).

3.4 Effect size and schools with different starting points

Expressed as a standardized effect size, the result for years with sustained use corresponds to approximately 0.11 standard deviations of school results, and the first-year result to 0.14 standard deviations of school results (Appendix Table A7). To illustrate the scale of these differences, an average difference of 0.30 grade points would be equivalent to a 2.5-point increase for approximately one in eight students, with the remaining students' results unchanged. The corresponding illustration for 0.24 grade points is approximately one in ten students. These are arithmetic equivalents, not estimates of how many individual students improved. A 2.5-point increase corresponds to one step between adjacent passing grades. Grouping schools by their Fail Rate before Magma, the Fail Rate was lower in years with sustained use in all three groups (Table 3).

Table 3. Fail Rate in years with sustained use, by Fail Rate before Magma

Fail Rate 2018/2019	Schools	Fail Rate without Magma	Reduction, percentage points	Relative reduction	p-value
Below 10%	1,048	9.0%	1.0	11%	0.08
10–20%	756	14.5%	1.3	9%	0.18
Above 20%	447	20.8%	1.7	8%	0.24

Reduction in failing students per 100 in years with sustained use (40+ active weeks), compared with the same school's years without Magma, estimated separately for each group, counting "~100" as 2.5 failing students. Each group contains fewer schools than the full sample, so the pooled estimate (1.2 fewer per 100, $p<0.01$) is the main result.

4 Discussion

Across the analyses, the most consistent finding concerns sustained use. Within the same schools, results were higher in years when Magma was used more weeks; the association held when schools were compared only within the same county and year, and it remained positive when only activity before May was counted. Schools with lighter use showed no statistically detectable difference from schools that never used Magma, and schools with high use showed no statistically detectable difference in the years before they started. The first-year and long-term comparisons point in the same direction.

The findings support an emphasis on sustained engagement when implementing Magma. Schools with sustained activity showed stronger mathematics results across several specifications, making regular use an important focus for further evaluation.

4.1 Limitations

This is an observational study, not a controlled experiment: schools were not randomly assigned to use Magma, so the results describe associations. The results should be read with the following in mind:

- Schools choose whether and how much to use Magma. School fixed effects and county-by-year controls account for stable differences between schools and for regional changes over time, but not for other changes at a school, such as staffing or teaching practices, that may coincide with sustained use.
- Each exam is taken by a different cohort of Year 6 students, and Skolverket publishes results at school level only, so the same students cannot be followed over time.
- Skolverket does not publish results based on few students and reports pass rates close to 100% as "~100".

5 Conclusion

This analysis matched Magma usage data with the published results of Swedish schools on the Year 6 national exam in mathematics from 2018/2019 to 2024/2025. Sustained use of Magma was associated with higher results and fewer failing students within the same schools, after accounting for differences between schools and for regional changes over time. Among schools starting in 2021/2022–2023/2024, high use was associated with stronger first-year results, and schools that used Magma from 2021/2022 onward moved from a higher to a lower Fail Rate than schools that never used Magma. Studies following individual students would add further detail to these findings.

A Supporting data

The tables below report the samples and the alternative specifications behind the results in this paper. All figures are produced from Skolverket's published results and Magma's usage data. Positive associations appear across multiple specifications, with the strongest evidence concerning sustained use.

Table A1. Data coverage by national exam

National exam	Schools with a result	Students with an exam grade	Schools with 30+ active weeks	National average, grade points
2018/19	2,714	107,877	9	12.00
2021/22	2,772	111,821	199	12.11
2022/23	2,790	114,249	463	11.53
2023/24	2,786	110,499	730	11.28
2024/25	2,821	112,407	1,015	11.44

The number of schools using Magma at least 30 active weeks in the school year of the exam grew from 9 in 2018/19 to 1,015 in 2024/25. National average: student-weighted average of municipal schools.

Table A2. First year of use, by start year

Start year	Exams compared	High use	Lighter use	Never used Magma	Difference, high use	p-value
Included in the main estimate (pooled: +0.30, $p<0.01$, 308 schools)						
2021/22	2018/19 → 2021/22	43	47	594	+0.32	0.27
2022/23	2021/22 → 2022/23	133	63	756	+0.38	0.01
2023/24	2022/23 → 2023/24	132	51	765	+0.21	0.21
Not included in the main estimate (continued use not yet observable)						
2024/25	2023/24 → 2024/25	141	65	781	−0.17	0.29

All three start years in the main estimate show a positive difference, between +0.21 and +0.38 grade points; individual start years have fewer schools, so the pooled estimate is the main result. Difference: change in grade points from the last exam before Magma to the first exam with Magma for schools with high use, relative to schools that never used Magma, controlling for the previous result. The main estimate pools the start years 2021/22 to 2023/24 (308 schools), for which continued use in the following years can be observed. Schools that started in 2024/25 are shown for completeness; their continued use cannot yet be observed.

Table A3. First year of use, alternative methods

Method	Difference, grade points	Std. error	p-value	High-use schools
Main estimate				
Previous result	+0.30	0.10	0.004	308
Alternative methods				
No controls	+0.29	0.10	0.006	308
Matched pairs	+0.36	0.14	0.012	308
Previous result and county	+0.18	0.11	0.089	308
No requirement of continued use after the first year	+0.28	0.10	0.006	328
Permutation test (500 random reassignments)	+0.30	–	0.004	308
Including the 2024/25 start year				
All start years including 2024/25, previous result	+0.15	0.09	0.095	449
All start years including 2024/25, previous result and county	+0.04	0.09	0.648	449
Before starting Magma				
Same schools, years before starting	0.00	0.13	0.98	–

Every method gives a positive difference. Before starting Magma, the same schools showed no statistically detectable difference from schools that never used Magma. Difference in the change in grade points between schools with high use and schools that never used Magma. Permutation test: share of 500 random reassignments of the high-use label, within start years, giving a difference at least as large as observed. Last row: change from 2018/19 to 2021/22 for the schools that started in 2022/23 or 2023/24, before they used Magma.

Table A4. First-year difference by definition of high use

	50+ problems per student	100+ problems per student	200+ problems per student
20+ active weeks	+0.23 (469)*	+0.25 (451)*	+0.28 (353)*
25+ active weeks	+0.29 (387)*	+0.28 (382)*	+0.29 (311)*
30+ active weeks	+0.31 (309)*	+0.30 (308)*	+0.31 (266)*
35+ active weeks	+0.24 (215)	+0.24 (215)	+0.24 (190)
40+ active weeks	+0.15 (72)	+0.15 (72)	+0.28 (65)

All fifteen definitions give a positive difference, and all nine definitions from 20 to 30 active weeks are statistically significant. Fewer schools reach 35 or 40 active weeks, which widens the margin of error. Grade points relative to schools that never used Magma, controlling for the previous result; number of high-use schools in parentheses. * $p<0.05$.

Table A5. Within-school comparison, sample

	Value
Schools with at least three exam results	2,791
School-years	12,904
School-years without Magma	8,730
School-years with under 30 active weeks	1,822
School-years with 30–39 active weeks	1,477
School-years with 40+ active weeks	875

Table A6. Fail Rate under alternative comparisons

Comparison	Magma schools	Never used Magma	Fewer failing students	p-value
2021/22 starters, published results 2018/19 and 2024/25	89	1,137	24%	–
As above, "~100" counted as 2.5 failing students	89	1,137	18%	–
2022/23 starters, same schools before and after the first year	190	727	20%	0.04
As above, "~100" counted as 2.5 failing students	190	727	15%	0.02
All schools, years with sustained use (40+ active weeks) vs. the same school's years without	513	–	9%	0.007

All comparisons show fewer failing students in schools using Magma, between 9% and 24%. Fewer failing students: reduction in the Fail Rate relative to schools that never used Magma, as a share of the Fail Rate before Magma (rows 1–4) or in the same school's years without Magma (row 5). Pass rates published as "~100" are counted as 100% unless stated. The last row counts them as 2.5 failing students.

Table A7. Standardized effect sizes

Result	Grade points	SD of student results	SD of school results
First year, high use	+0.30	0.05	0.14
Years with sustained use (40+ active weeks)	+0.24	0.04	0.11

SD of school results: 2.17 grade points (unweighted standard deviation across municipal schools, 2024/25). SD of student results: approximately 5.48 grade points, calculated from an assumed national grade distribution rather than student-level data.

Table A8. Within-school Fail Rate under both treatments of "~100"

Active weeks	**2.5 failing, school + year**	**As published, school + year**	**2.5 failing, county × year**	**As published, county × year**
Under 30 active weeks	−0.27	−0.45	−0.25	−0.43
	($p = 0.32$)	($p = 0.15$)	($p = 0.36$)	($p = 0.18$)
30–39 active weeks	−0.52	−0.63	−0.43	−0.52
	($p = 0.11$)	($p = 0.10$)	($p = 0.20$)	($p = 0.18$)
40+ active weeks	−1.19 ($p<0.01$)	−1.20	−1.01	−0.97
		($p = 0.02$)	($p = 0.03$)	($p = 0.07$)

All twelve estimates show fewer failing students in years with Magma, and the reduction is largest with 40+ active weeks under both treatments. Change in failing students per 100 compared with the same school's years without Magma. "2.5 failing": pass rates published as "~100" counted as 2.5 failing students (used in the main text). "As published": counted as 100% passing. Standard errors clustered by school.