Kami in Moreno Valley Unified School District

November 2023

This report leverages product usage data and school district data to explore the association between Kami usage and student outcomes.

Principal Findings

Empirical Education conducted a study to evaluate the impact of Kami on student achievement in reading in Moreno Valley Unified School District in the 2022–23 school year.

- There was a significant positive association between the use of Kami and student outcomes on the STAR Reading test. Daily usage of Kami during the school year is associated with a 4 percentile test score gain.
- The impact was greater for English language learners (ELL) compared to non-ELL students.
- There were no significant differences between students of different races/ethnicities.

BACKGROUND

This study of usage and effectiveness of Kami is based on student data from Moreno Valley Unified School District and student-level application usage data from Kami from the 2022–23 school year. Students in all grades across the district used Kami to create, share, annotate, and collaborate on documents across subject areas. The main research question is whether usage of Kami has a positive impact on student achievement in reading.

STUDY DESIGN

This study follows a correlational design aimed at establishing statistical associations between product usage metrics and student outcomes. Unlike experimental and quasi-experimental studies, it does not compare users to similar non-users. It focuses entirely on product users and the differences in outcomes among them that can be attributed to the usage, making appropriate adjustments for differences in users' individual and class characteristics and pretest scores.

The results of such a study are used to predict potential outcomes at some possible level of usage. In this study, that level is consistent daily use throughout the school year. We compare those potential outcomes to imputed outcomes for non-users. Positive results should be reviewed as showing potential (promise of effectiveness) rather than proving effectiveness, because there is a possibility that more capable students and/or more effective teachers choose to use the application more actively. Therefore, the observed positive effect may be due to underlying user qualities and not the application. This *self-selection bias* is a typical issue in non-experimental studies. Analysis of determinants of usage is undertaken to identify and reduce possible bias. However, potential self-selection due to factors not reflected in the data cannot be adjusted for.

DATA

Data collected for this study consisted of 20,680 individual student records containing student demographics, school and teacher identifiers, pretest and posttest scores (Fall 2022 and Spring 2023 respectively) from the STAR Reading assessment, and multiple metrics of Kami usage. Student-level data were further aggregated into class and school usage metrics, including class and school test score averages, proportions of various student groups, and average use by school. Class was defined as all students associated with the same teacher identifier but not necessarily in the same classroom (section).

A subset of student records (about 75% of the total) with both pretest and posttest scores available were included in the sample for analysis. Additional reduction of the analytic sample was warranted by the need to mitigate potential selection bias. Preliminary analyses showed that Kami usage, as measured by user active days, was negatively correlated with students' disability status. Removing special education classes (classes with 100% students with disability) from the sample eliminated that association and therefore reduced potential selection bias. This adjustment reduced the sample by approximately two percent.

Parameters of the final analytic sample are presented in Tables 1-2.

TABLE 1. SAMPLE SIZES

Category	All data	Analytic sample
Schools	38	36
Classes	612	517
Students	20,680	15,255

TABLE 2. CHARACTERISTICS OF ANALYTIC SAMPLE

Category	% Total
Free/reduced-price lunch	72.4
English language learner	19.1
Special Education	11.2
White	7.1
Black	11.8
Hispanic	75.6
Other ethnicities	5.5

The distribution of students by grade in the sample and average usage statistics are presented in Table 3. These results do not reveal any apparent patterns (such as increasing or decreasing usage with grade) suggesting that Kami can be widely adopted in classes of any grade level.

TABLE 3. USAGE BY GRADE

Grade	Students	Average Usage Days	Average Events
Grade 2	823	32	1,895
Grade 3	1,473	20	938
Grade 4	1,530	32	1,577
Grade 5	1,879	36	1,528
Grade 6	1,779	47	1,698
Grade 7	1,698	44	1,199
Grade 8	1,482	56	1,926
Grade 9	1,681	40	1,246
Grade 10	1,532	50	1,861
Grade 11	1,378	38	1,271

ANALYSIS

The analysis of association between Kami usage and student outcomes was performed using a hierarchical mixed-effects linear regression model. The Spring 2023 scores on the STAR Reading assessment were the outcome variable, and student and class characteristics and pretest scores were covariates. Pretest and posttest scores were normalized by grade putting all scores on the same scale. The model makes adjustments for students' clustering in classes and schools.

The model estimates the association between student test score gains and a 1-unit increase in Kami usage (one active day). We use these estimates to project the differences in outcomes between non-users (zero active days) and users with average usage for the sample. We also projected this for 'optimal' users: those who use Kami daily throughout a 180-day long school year. The results are presented as percentile gains for a hypothetical Kami user who would score at the 50th percentile if they were enrolled in a class not using Kami. These estimates are presented in Table 4 for all students and for student groups.

Effects for student groups were estimated by including interaction terms in the model, allowing us to identify differences in the association between Kami usage and student characteristics. The effects reported here estimate the potential differences in outcomes between two 'average students' who only differ in one characteristic (e.g. ELL status) but are otherwise identical.

RESULTS

We found evidence that usage of Kami, as measured by active days, is positively associated with student outcomes (Spring 2023 STAR Reading). The average student used Kami on 40 distinct days - this amount of usage is associated with a 1 percentile point test score gain. We estimate that optimal daily usage of Kami would be associated with an impact equivalent to a 4 percentile point increase on the STAR assessment. The level of confidence we have in these estimates is high.

Kami usage is positively associated with student outcomes in all student groups. However, the differential association between usage and outcomes is established only for ELL students vs non-ELL students: ELL students with optimal Kami usage can be expected to achieve an 8 percentile test score gain. ELL students, therefore, appear to benefit from Kami usage the most of all student groups. There were no statistically significant differences in the impact of Kami according to the student's race/ethnicity or across grade levels.

TABLE 4. STUDENT RESULTS OVERALL AND BY GROUP

Category	Effect for average use (percentile)	Effect for daily use (percentile)	Significant differential
All	1	4	
Female	1	5	No
Male	1	3	No
ELL	2	8	Yes
Non-ELL	1	3	Yes
Disability	2	7	No
Non-disability	1	4	No
FRPL	1	5	No
Non-FRPL	0	2	No

In addition to the analyses reported above, we performed an exploratory analysis of the association between various usage metrics and student outcomes. In these models, the single usage metric (active days) was replaced with an array of usage metrics including counts of different types of events and characteristics of sessions. The model was estimated iteratively with the elimination of least significant terms until only a few significant metrics remained in the final model. The remaining metrics are shown in Table 5.

TABLE 5. ASSOCIATION OF VARIOUS METRICS WITH STUDENT OUTCOMES

Metric	Estimate	Std. Error	*p* value
Documents loaded	-0.00051	0.00013	<.001
Annotations modified	0.00005	0.00001	<.001
Optical character recognition success	0.0059	0.0021	<.01
LMS event	0.0001	0.00005	0.02
Proportion of longer sessions (>10 minutes)	0.088	0.029	<.01

These results should be interpreted with caution because it is impossible to single out the effect of one type of event from others given the strong correlations among all metrics. In addition, the scales and variability of each metric are different.

One noteworthy result is the positive effect of longer sessions. The difference between two hypothetical students, one of whom only briefly interacts with Kami (no matter how frequently) and another who spends 10 minutes or more each time, is equivalent to 4 test score percentiles (which corresponds to the difference between zero and the 0.088 contribution of longer sessions). The negative effect of the number of documents loaded vs. the positive effect of annotations modified may be indicative of the beneficial effect of the greater focus on a smaller number of documents.

CONCLUSION

Results of this study present strong evidence of promise that Kami can improve student outcomes across student groups and grade levels. The study sample is large enough to obtain statistically significant results for student groups. However, the generalizability of these results is somewhat limited by the demographics of participating students, three quarters of which belong to one ethnic group. In the interpretation of the results, it is important to remember that this is a non-experimental study, with no defined treatment and control groups, and that the reported usage effects are projections based on correlational results.

Appendix. Technical Details

This study relied on an earlier analysis of Kami usage that revealed a straightforward pattern in usage: active users tend to be engaged simultaneously in many events. As a result, various usage metrics are correlated with each other. In particular, "active days" exhibits strong correlations with all other metrics and can therefore be chosen as an indicator of usage overall.

A majority of students used Kami at least once weekly, and this group generated the most events in the system (see table below).

TABLE 6. FREQUENCY OF USAGE

Frequency of use	Students	Average usage days	Average events per user	Total events (millions)
Low (once or month or less, 1-9 days a year)	3,701	4	78	0.3
Moderate (up to once a week, 10-36 days a year)	4,650	22	633	2.9
High (more than once a week, >36 days a year)	6,904	71	2,839	19.6

Analysis of variance showed that most variation in Kami usage is at the school and class level. Differences in usage frequency among students in any particular class account for less than one fifth of the total variance. This is consistent with Kami being a collaborative product.

TABLE 7. VARIANCE OF USAGE BY LEVEL

Student	18.8%
Class	32.4%
School	48.7%

Multilevel regression analysis of potential determinants of usage revealed that the variation was mostly due to random (unobserved) factors with one exception: class-average active days were negatively correlated to the proportion of students with a disability in the class. Removing special education classes (classes with 100% students with a disability) was sufficient to eliminate that association.

The tables below present detailed results of the analysis of the associations between Kami usage and student outcomes. The main results are reported as the “effect per unit (one day) of usage.” Since the test scores were normalized prior to the analysis, these estimates are expressed in the units of standard normal distribution (e.g. an effect equal to 1 would mean the score gain equal to one standard deviation of test scores). Multiplying this effect by the actual average days (40) or “optimal” usage of 180 days (a typical school year) put these results in perspective, demonstrating the potential gain for users vs. non-users. Using 180 days as the benchmark does not imply that this level of usage is practical, but only sets the absolute upper limit to potential gains. The '*p* value' is the measure of the precision of the results or the strength of evidence that the effect in question is statistically different from zero. Conventional interpretation is that a *p* value of .05 or less signifies strong evidence, and *p* values above .05 but less than .20 provide limited evidence. Higher *p* values imply that our results provide no reliable information about the association of the usage and outcomes, since the probability that the true effect is zero—or even has an opposite sign—is too high. Higher *p* values (lower precision of the results) are typical when the student group is small. High or low effect estimates should be ignored when the *p* values are greater than .20.

TABLE 8. STUDENT RESULTS OVERALL, BY RACE/ETHNICITY, AND BY STUDENT GROUP

Category	Effect per unit (usage day)	Standard error	*p* value	Effect of daily use	Effect of daily use (percentiles)	Effect of average use	Effect of average use (percentiles)
All	0.00055	0.00018	<.01	0.099	4	0.022	1
Asian	-0.00039	0.00080	.63	-0.07	-3	-0.016	-1
Black	0.00072	0.00042	.09	0.13	5	0.029	1
Hispanic	0.00054	0.00020	<.01	0.098	4	0.022	1
Native American	0.0049	0.0034	.16	0.88	31	0.19	8
Pacific Islander	0.00032	0.00090	.72	0.057	2	0.013	1
Unknown	0.00038	0.00160	.81	0.068	3	0.015	1
White	0.00086	0.0005	.11	0.16	6	0.035	1
Female	0.00065	0.00022	<.01	0.12	5	0.026	1
Male	0.00045	0.00023	.047	0.081	3	0.018	1
ELL	0.0011	0.00033	<.01	0.199	8	0.044	2
Non-ELL	0.00041	0.00019	.037	0.073	3	0.016	1
Disability	0.00105	0.00043	.015	0.188	7	0.042	2
Non-disability	0.00049	0.00019	<.01	0.089	4	0.019	1
FRPL	0.00068	0.00020	<.001	0.122	5	0.027	1
Non-FRPL	0.00022	0.00028	.43	0.040	2	0.0088	0