

Goalbook Pathways in Lima City Schools

August 2019

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

Introduction and Principal Findings

The Evidentially, Inc. research team identified Lima City, OH as a potential research site by conducting an analysis of usage of the Goalbook Pathways product in schools/districts in the US during the 2017-18 school year. After entering a data-sharing agreement with the district, we obtained student ELA outcomes from the Scholastic Reading Inventory (SRI), a formative assessment administered in the fall, winter, and spring to measure reading comprehension, as well as student demographic characteristics for students in grades 3 through 8. We matched these to product usage data collected from the Goalbook Pathways product. This study explores patterns of usage of Goalbook Pathways by teachers and measures the association between the product usage and SRI outcomes in Spring 2018 for students in upper elementary and middle grades. The principal findings are:

- The count of “Viewed Unpacked Standard” action is the single best predictor of student outcomes. Using this metric, Goalbook Pathways shows a strong evidence of promise.
- The strongest association with student outcomes are observed in eighth grade and for male and for black students.
- Teachers, rather than school or district leaders or students, primarily made their own decision to use Goalbook Pathways.
- There is a wide variation in Goalbook Pathways usage among teachers and across schools: the average days of usage range between 1.5 and 20 days across schools and vary between 0 and 73 days for individual teachers.

Usage: Correlations among Metrics

The first step in this study was to identify metrics of product usage. We used a statistical technique called “principal component analysis” (PCA) that reduces large sets of variables into smaller ones (principal components) that still capture most of the information in the larger set. Usage of Goalbook Pathways is characterized by a straightforward pattern: active users tend to be engaged simultaneously in many events. There are statistically significant positive correlations among all events, as shown in Table 1, and the first component in the PCA accounts for most (63%) of the total variation among usage.

The first principal component can be used as the overall "usage score". The usage score is highly correlated with usage days and total events, so that all three composite metrics adequately reflect the intensity of use of Goalbook Pathways by teachers.

TABLE 1. CORRELATIONS BETWEEN EVENTS

Event	Shown Item Preview	Viewed Collection Detail	Viewed Passage Detail	Viewed Project Detail	Viewed Resource Detail	Viewed Standard Resources	Viewed Strategy Detail	Viewed Unpacked Standard	Viewed Your Collections	Viewed Your Favorites
Shown Item Preview	*	0.23	0.70	0.16	0.35	0.46	0.27	0.32	0.23	0.52
Viewed Collection Detail	0.23	*	0.35	0.17	0.19	0.30	0.28	0.35	1.00	0.13
Viewed Passage Detail	0.70	0.35	*	0.34	0.30	0.45	0.28	0.30	0.35	0.18
Viewed Project Detail	0.16	0.17	0.34	*	0.82	0.80	0.81	0.51	0.17	0.60
Viewed Resource Detail	0.35	0.19	0.30	0.82	*	0.80	0.76	0.56	0.19	0.83
Viewed Standard Resources	0.46	0.30	0.45	0.80	0.80	*	0.76	0.76	0.30	0.70
Viewed Strategy Detail	0.27	0.28	0.28	0.81	0.76	0.76	*	0.54	0.28	0.63
Viewed Unpacked Standard	0.32	0.35	0.30	0.51	0.56	0.76	0.54	*	0.35	0.44
Viewed Your Collections	0.23	1.00	0.35	0.17	0.19	0.30	0.28	0.35	*	0.13
Viewed Your Favorites	0.52	0.13	0.18	0.60	0.83	0.70	0.63	0.44	0.13	*

Note. The numbers in the table are Pearson correlation coefficients, r . The correlation coefficient is a value between -1 and 1 that describes 1) the strength of the association between two variables, and 2) the direction of the association. The closer that $|r|$ is to 1, the stronger the two variables are associated with each other. If r is positive, thus describing a positive association, then as one variable increases in value, typically the other variable increases. If r is negative, then as one variable increases in value, typically the other variable decreases. If $r = 0$, then there is no association between the two variables.

Usage: Types of Events

As shown in Table 2, two types of events - "Viewed Standard Resources" and "Showed Item Preview" - stand out as the most widely used and most correlated with overall usage score (first principal component). "Viewed Unpacked Standard" follows closely in terms of correlation but is less frequently used.

TABLE 2. MEANS OF EVENTS AND CORRELATIONS WITH USAGE SCORE

Events	Mean	Correlation with Usage Score
Viewed Standard Resources	27.9	0.61
Showed Item Preview	19.0	0.60
Viewed Unpacked Standard	5.8	0.40
Viewed Project Detail	3.2	0.20
Viewed Resource Detail	3.2	0.18
Viewed Strategy Detail	3.8	0.15
Viewed Your Favorites	2.3	0.09
Viewed Passage Detail	1.7	0.05
Viewed Your Collections	1.0	0.01
Viewed Collection Detail	1.0	0.00

Note. Means calculated for the sample that excludes teacher with no recorded usage of Goalbook Pathways. Usage Score is the first principal component.

Usage: Teachers are the Primary Source of Variation

There is substantial variation in average usage across schools (Table 3).

TABLE 3. VARIATION IN USAGE BETWEEN SCHOOLS

School	Average usage days per school	Average events per school
Freedom	20.29	99.9
South	19.14	69.09
West	17.9	92.3
Independence	14.3	61.11
North	12.86	47.74
Unity	12.82	69.97
Liberty	3.87	65.91
Heritage	1.53	6.22

Note. Usage metrics' averages are class size-weighted, i.e. larger classes contribute more to the average. Schools are ordered by average number of usage days.

However, much more variation in usage days is among teachers within schools than between schools - 78% vs. 22% - suggesting that teachers mostly independently choose their level of engagement with Goalbook Pathways, as opposed to suggested or prescribed use by school leadership.

Teachers can be divided into three groups by the frequency of usage: non-users, occasional users, and active/frequent users. Of the two last groups, which are about equally sized, active/frequent users account for most usage (Table 4).

TABLE 4. VARIATION IN USAGE BETWEEN TEACHERS

Frequency of use	Teachers	Average usage days	Average events per user	Total events
Never	7	0	0	0
Occasionally (1-8 times a year)	27	2.6	13	442
Active/frequent (9 or more - at least monthly on average)	23	23.3	109	2,500

There is a general tendency for the more active users of Goalbook Pathways to use the product more often and to use it more intensively during each session. However, there is a lot of variation in the temporal patterns of use, with the greatest variability of intensity of each session among occasional users, as illustrated by the following figure (Figure 1). Each dot represents a teacher; occasional users (1-8 usage days per year) have a wider range in “events per day” than active/frequent users (>9 usage days per year). Goalbook developers identified optimal weekly use of Pathways as an appropriate goal, the equivalent of 36 usage days in a school year, and Table 5 shows the full distribution of usage days by teacher.

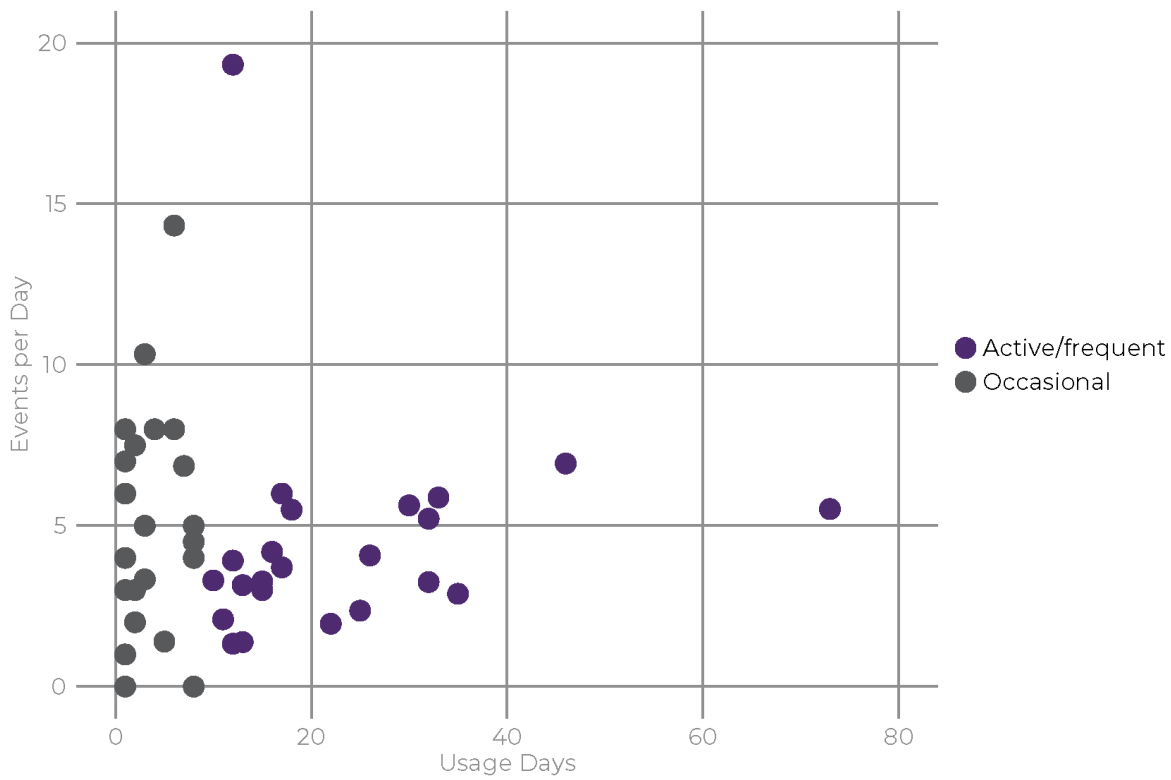


FIGURE 1. SCATTERPLOT OF USAGE DAYS VERSUS EVENTS PER DAY

TABLE 5. DISTRIBUTION OF USAGE DAYS BY TEACHER

Frequency of use (days)	0	1	2-4	5-7	8-14	15-24	25-35	36 or more
No. of Teachers	7	12	7	4	11	7	7	2

Usage: Classroom Characteristics

There are some important associations between class characteristics and Goalbook Pathways usage. The strongest influences are proportions of male students and students who transferred during the school year. A one percentage-point difference in the proportion of transfer students is associated with 5 percent difference in Goalbook Pathways usage, whereas one percentage-point difference in the proportion of male students is associated with 3 percent difference in Goalbook Pathways usage. However less than one quarter of total variation in Goalbook Pathways usage is associated with class characteristics. Most differences in Goalbook Pathways usage across classrooms are due to teachers' choices unrelated to class characteristics.

Effect of Goalbook Pathways Usage on Student Outcomes

There is limited evidence that active usage of Goalbook, as measured by aggregate metrics (i.e., usage days, total events, or usage score), is positively associated with student outcomes (Spring 2018 SRI). If usage days is used as the single measure of use, positive effects can be identified in grades 3 and 8 and for black students. The level of confidence we have in these estimates is moderate (see Technical Details). Using total events or the usage score as the usage metrics produces a similar result but with a lower level of confidence. These results may be influenced by a small number of teachers with very high total usage.

EFFECT OF "VIEWED UNPACKED STANDARDS"

Analysis of outcomes with counts of individual types of action as usage metrics established that "Viewed Unpacked Standard" is the single best predictor of student outcomes. Using the count of "Viewed Unpacked Standard" as the usage metric produces results that are largely consistent with the results obtained for usage days but the statistical results give us greater confidence. Potential strong positive effects are identified for black and male students and in grades 4 and 8. We have strong confidence in the average effect, as well as for effects on non-English Language Learners (ELL) and non-Special Education students.

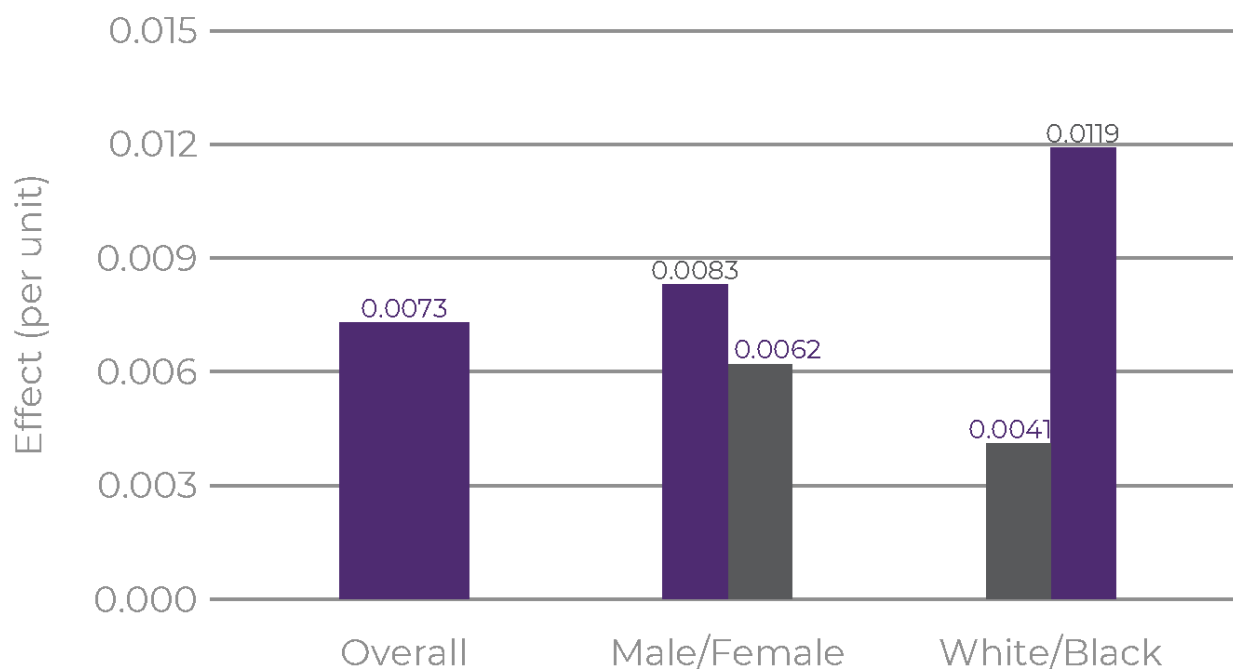


FIGURE 2. OVERALL EFFECT OF "VIEWED UNPACKED STANDARDS" ON SRI STUDENT OUTCOMES

Note. Purple bars indicate $p < .05$.

With the Goalbook developers' suggested optimal usage—weekly throughout the year or 36 usage days—and assuming no diminishing returns to "viewing unpacked standards", the average positive effect could be as large 10 percentiles and up to 16 percentiles for black students. A one-percentile test score gain is associated on average with three counts of the "Viewed Unpacked Standard" action.

These estimates do not show that Goalbook use by teachers has a strong direct positive effect on student outcomes but they do show promise of positive effects. More information on the interpretation of these results can be found in 'Technical details' section.

Disclaimer: It must be taken into account that these results are obtained using a relatively small sample in a particular school district and may not necessarily be replicable elsewhere.

Technical Details

This study of usage and effectiveness of Goalbook Pathways is based on student data from Lima City Schools and teacher-level application usage data from Goalbook from the 2017-18 school year. Student data included demographics, school and teacher identifiers, pretest score (Fall 2017), and test score (Spring 2018) from the Scholastic Reading Inventory (SRI). Student-level data were further aggregated into class and school usage metrics, including class and school test score averages, proportions of various student subgroups, and average use by school. Class was defined as all students with the same teacher IDs in the same grade, who may not necessarily be in the same classroom (section). Teacher data included in this study consisted of the counts of actions (events) of ten selected types, listed in Table 1 in 'Usage' section of this report.

Three aggregate usage metrics were created for use in the analysis:

1. Usage days – the total number of days in the 2017-18 school, on which any application usage activity by a given teacher is recorded
2. Total events – a sum of all activity counts in a year
3. Usage score – the first principal component from the PCA

PCA-based usage scores was of limited utility for this study because they are not informative for teachers without any usage, and such teachers constituted a substantial proportion of the sample. Most analyses presented in this report were performed with usage days and total events.

STUDY DESIGN

This study follows a correlational design aimed at establishing statistical association between product usage metrics and student outcomes. Unlike experimental and quasi-experimental studies, it does not compare users to similar non-users, but focuses entirely on product users and 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, daily use throughout the school year) and compare those to imputed outcomes for zero usage. Positive results

should be taken as showing potential promise of effectiveness because there is a possibility that more capable users (highly effective teachers) choose to use the application more actively and therefore the observed positive results may be due to underlying user qualities and not the application effect per se. This is a typical issue in correlational design. In this study, its interference is likely to be limited because the effectiveness analysis is performed at the student level while the usage is determined at the class level. At the same time, we find that only about one quarter of variation in the usage is associated with class characteristics so Goalbook Pathways usage is unlikely to be determined by those. In future studies it would help to have data on teacher characteristics such as years of experience and teacher effectiveness rankings. In this study we were unable to establish the extent to which usage may be determined by teacher quality.

ANALYTIC SAMPLE

The analytic sample in this study excluded classes (as defined earlier) that were very small (four or fewer students consisting mostly of special-education students), where no pretest or test data were available. Teachers who had only such small classes did not appear consequently in the analytic sample. The final parameters of the analytic sample are presented in Tables 6-9

TABLE 6. SAMPLE SIZES

Category	All data	Analytic sample
Schools	8	8
Teachers	57	52
Classes	73	60
Students	1711	1680

TABLE 7. CHARACTERISTICS OF STUDY SAMPLE (STUDENTS)

Category	% Total
FRPL	100
ELL	0.5
Special Education	17.4
White	36.6

Category	% Total
Black	39.9
Hispanic	4.5
Class Size	42
Transfer	10
Grade 3	18.3
Grade 4	15.7
Grade 5	15.8
Grade 6	15.6
Grade 7	17.0
Grade 8	17.6

TABLE 8. CHARACTERISTICS OF STUDY TEACHERS BY GRADE

Grade	Teachers	Average Usage Days	Average Events
Grade 3	15	10.5	54
Grade 4	14	10.4	42
Grade 5	7	12.9	56
Grade 6	6	11.5	71
Grade 7	10	25.3	149
Grade 8	8	5.4	21

TABLE 9. DISTRIBUTION OF EVENTS PER USAGE DAY

Events Per Day	Usage Days
1	141
2	83
3	60
4	57
5	39
6	28
7	28
8	13

Events Per Day	Usage Days
9	15
10	19
11-15	31
16-20	10
21 or more	22

ANALYSIS

The analysis was performed using a linear regression model with the Spring 2018 scores on the SRI as the outcome variable, and student and class characteristics and pretest scores as covariates. The model makes an adjustment for students clustering in classes. Subgroup effects were estimated from moderator analyses whereby the single usage term is replaced with interaction terms. Subgroup effects, reported here, are adjusted for the differences in student characteristics, and thus, they estimate the potential difference in outcomes between two 'average students' who only differ in one characteristic (e.g. ELL status) but are identical otherwise. The tables below (Tables 10 and 11) report the results in the first column as the “effect per one day of usage” 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). The last column converts this into percentile gain for weekly usage (36 days during the school year) for a hypothetical student who would score at the 50th percentile if he/she were studying with a teacher not using Goalbook Pathways. This does not imply that this level of usage is practical, but only sets the absolute upper limit to potential gains. The percentile gain field includes "NA" if the estimate is not statistically significant. 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 effect of usage on 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 subgroup of students is small and does not necessarily mean that there is no effect.

Table 10 presents the results with usage days as the aggregate measure of application usage while Table 11 presents the results that rely on “Viewed Unpacked Standard”, which was identified as the single best predictor of student outcomes. It must be emphasized that the predictive power of this type of user action does not necessarily mean that it is the most important per se but it is associated with effective patterns of application usage.

TABLE 10. DETAILED RESULTS WITH USAGE DAYS

Category	n	Effect per unit (usage day)	Standard error	p value	Potential effect of optimal weekly use (36 days)
Average Effect	1680	0.0007	0.0015	.65	NA
Male	904	0.0008	0.0019	.66	NA
Female	776	0.0005	0.0020	.79	NA
White	615	-0.0013	0.0023	.99	NA
Black	671	0.0031	0.0021	.13	4
Hispanic	75	-0.0008	0.0053	.99	NA
ELL	9	-0.0119	0.0185	.99	NA
Not ELL	1671	0.0008	0.0015	.61	NA
Special Education	293	-0.0042	0.0038	.99	NA
Not Special Education	1387	0.0014	0.0016	.37	NA
Third grade	308	0.0039	0.0027	.15	5
Fourth grade	264	0.0017	0.0033	.62	NA
Fifth grade	265	0.0033	0.0041	.43	NA
Sixth grade	262	-0.0106	0.0178	.99	NA
Seventh grade	286	-0.0051	0.0026	.99	NA
Eighth grade	295	0.0131	0.0073	.07	18

TABLE 11. DETAILED RESULTS WITH “VIEWED UNPACKED STANDARD”

Category	n	Effect per unit ("Viewed Unpacked Standards")	Standard error	p value	Potential effect of optimal weekly use (36 days)
Average Effect	1680	0.0073	0.0028	.01	10
Male	904	0.0083	0.0036	.02	11
Female	776	0.0062	0.0039	.11	8
White	615	0.0041	0.0043	.34	NA
Black	671	0.0119	0.0043	.01	16
Hispanic	75	-0.0010	0.0106	.99	NA
ELL	9	-0.0232	0.0451	.99	NA
Not ELL	1671	0.0075	0.0028	.01	10
Special Education	293	-0.0012	0.0074	.99	NA
Not Special Education	1387	0.0084	0.0030	.01	11
Third grade	308	0.0060	0.0056	.28	NA
Fourth grade	264	0.0122	0.0046	.01	16
Fifth grade	265	0.0109	0.0085	.20	NA
Sixth grade	262	0.0055	0.0112	.63	NA
Seventh grade	286	-0.0064	0.0067	.99	NA
Eighth grade	295	0.0209	0.0124	.09	27