Exploratory Analysis of Students’ Mathematics Achievement After Using Freckle

Written in partnership with WestEd
## Table of Contents

Abstract .............................................................................................................................................. 1  
Executive Summary .......................................................................................................................... 2  
Introduction ...................................................................................................................................... 3  
Method ............................................................................................................................................. 4  
  Participants ..................................................................................................................................... 4  
  Data Analysis ................................................................................................................................. 6  
Results .............................................................................................................................................. 6  
  Baseline Equivalence ..................................................................................................................... 6  
  Post-Score Analysis ....................................................................................................................... 7  
Discussion ......................................................................................................................................... 8  
  Study Limitations ........................................................................................................................ 8  
  Conclusions ................................................................................................................................. 8  
References ......................................................................................................................................... 9  
Appendix A ........................................................................................................................................ 10  
Appendix B ....................................................................................................................................... 13  

---

![Freckle](Freckle.png)

---

*WestEd* WestEd.org
List of Figures

Figure 1. Sample Report Card data view available for teachers in Freckle Math..........................3
Figure 2. Means for students’ beginning and end of year MAP scores by grade and condition........5
Figure 3. Covariate-adjusted means for each group’s end of year MAP scores..................................7

List of Tables

Table 1. School Demographic Information.......................................................................................4
Table 2. Teacher and Student Counts by Grade Included in Final Analytic Sample.........................5
Table 3. Growth norms observed in MAP scores throughout the year....................................................6
Table 4. Random Effects....................................................................................................................11
Table 5. Fixed Effects........................................................................................................................11
ABSTRACT

WestEd—a nonpartisan, nonprofit research, development, and service agency—partnered with Freckle to explore whether students who use Freckle exhibit different mathematics achievement outcomes than students who do not use Freckle. A rural school shared data on Freckle adoption and student scores on the Northwest Evaluation Association’s (NWEA) Measure of Academic Progress (MAP) in mathematics with Freckle. WestEd used these data to conduct an independent analysis to determine whether or not Freckle use was associated with higher NWEA MAP test scores. Results are consistent with the hypothesis that using Freckle in the classroom results in higher achievement in mathematics. Limitations to the study and future research directions are also discussed.
EXECUTIVE SUMMARY

Freckle is a computerized math and reading comprehension program that provides adaptive feedback and practice opportunities for K - 8 students. Freckle is hypothesized to improve students’ reading and mathematics outcomes in education. In the present study, WestEd explored the potential of Freckle to support students’ mathematics achievement in Kindergarten, Grade 1, and Grade 2 classrooms.

Study participants consisted of 466 Kindergarten, Grade 1, and Grade 2 students and 25 teachers in a rural elementary school during the 2014-2015 academic year. Teachers chose to use Freckle or chose not to use Freckle during the course of the school year. Approximately half of the teachers and their associated students (14 teachers, 263 students) used Freckle in their classrooms, whereas the other half (11 teachers, 203 students) participated in business-as-usual classroom activities.

This study explores whether results are consistent with the predictions that Freckle supports student learning in mathematics. Towards this end, students’ end of year mathematics scores on the Measure of Academic Progress (MAP) assessment was collected for each student. Preliminary analyses indicated that Freckle students exhibited higher beginning of year MAP scores than the comparison students. However, an analysis that attempted to control for beginning of year MAP score differences as well as grade level indicated that students in the Freckle condition scored approximately 4.6 points higher on their end of year MAP scores relative to students who did not use Freckle, a statistically significant difference ($p < .001$).

The results of this exploratory study are consistent with the hypothesis that Freckle positively impacts students’ mathematics achievement in K - 2 Grades. However, because teachers were not randomly assigned to use Freckle, the results should not be used to claim that Freckle caused higher student achievement. In addition, the study relies on a data set with limited information: Teachers chose to use Freckle or not for reasons unknown to the study team. In addition, potentially important demographic information on teachers and students are missing, thereby limiting the generalizability of the findings. Future studies can be conducted that utilize random assignment to groups and that collect information on factors that may potentially influence the effectiveness of Freckle. Such studies will increase confidence that Freckle use leads to improved mathematics outcomes.
INTRODUCTION

Freckle is an adaptive math and reading comprehension program that ensures every student receives high-quality, individualized math and reading practice. By addressing student weaknesses and building off of strengths, Freckle is designed to increase confidence and promote growth in all students, regardless of their ability level.

When starting on Freckle, students take a diagnostic to determine what they already know and areas that they could use more practice in. Then, as they continue practicing, the program adapts to provide individualized instruction tailored for each student’s skill level.

Freckle is designed to be used multiple times a week in the classroom to build students’ math fluency, and provides teachers data for each student such as the report shown below:

![Figure 1. Sample Report Card data view available for teachers in Freckle Math.](image)

This study addressed the question: Do students with teachers who used Freckle exhibit higher end of year mathematics scores than students from teachers who did not use Freckle?

Towards this end, WestEd, a nonprofit research, development, and service organization, analyzed data that was provided by the participating school, which included end of year mathematics scores on the Measure of Academic Progress (MAP) assessment for all Kindergarten, Grade 1, and Grade 2 students at the school. The MAP assessment is administered by the Northwest Evaluation Association (NWEA) and is an adaptive measure of students’ mathematics achievement. It was predicted that if Freckle supports students’ mathematics achievement, students of teachers who used Freckle would exhibit higher end of year scores on the MAP assessment relative to students of teachers who did not use Freckle.
METHOD

This study is not a randomized control trial study (i.e., students were not randomly assigned to conditions). Because of many threats to internal validity under this design, this study does not attempt to make causal claims on the impact of Freckle in student scores. Rather, it explores whether the results are consistent with the hypothesized predictions related to the impact of Freckle on student achievement.

PARTICIPANTS

Participants in this study were from a rural, public elementary school in the Southern United States (see Table 1). Participants consisted of 25 teachers (14 teachers in the Freckle group and 11 teachers in the Comparison group) and 466 students1 (263 students in the Freckle group and 203 students in the Comparison group). The 25 teachers comprised all of the teachers of Kindergarten, Grade 1, and Grade 2 at the participating school.

Some of the teachers used Freckle with their students whereas some of the teachers did not use Freckle (henceforth, these groups are referred to as the Freckle and Comparison groups, respectively). The school did not provide information on teachers’ reasons for opting to use Freckle. Students in both groups took the MAP Math assessment at the beginning and end of the 2014-2015 academic year (i.e., the Fall and Spring, respectively).

Attrition. Thirty-five students (16 students from the Freckle condition and 19 students from the Comparison condition) did not have complete pre and post-test data and therefore are excluded from all analyses. The complete analytic sample then consisted of 431 students (247 in the Freckle condition and 184 students in the Comparison condition) (see Table 2).

Table 1. School Demographic Information

<table>
<thead>
<tr>
<th>Type</th>
<th>% Free/Reduced Lunch</th>
<th>Enrollment</th>
<th>% White</th>
<th>% Black</th>
<th>% Hispanic</th>
<th>% Asian/PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public, Title I</td>
<td>69%</td>
<td>572</td>
<td>72%</td>
<td>21%</td>
<td>2.6%</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

1 One student id appeared in two classrooms in the datasets provided and is not considered in these analyses.
Table 2. Teacher and Student Counts by Grade Included in Final Analytic Sample

<table>
<thead>
<tr>
<th></th>
<th>Kindergarten</th>
<th>First Grade</th>
<th>Second Grade</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Students</td>
<td>Teachers</td>
<td>Students</td>
<td>Teachers</td>
</tr>
<tr>
<td>Freckle</td>
<td>66</td>
<td>4</td>
<td>86</td>
<td>5</td>
</tr>
<tr>
<td>Comparison</td>
<td>63</td>
<td>4</td>
<td>82</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>8</td>
<td>168</td>
<td>10</td>
</tr>
</tbody>
</table>

The average beginning and end of year MAP scores by grade and condition are provided in Figure 2.

Figure 2. Means for students’ beginning and end of year MAP scores by grade and condition.
Table 3 shows the observed and expected growth for students for each participating grade level and condition. MAP growth norms are based on the beginning to end of year growth on the MAP mathematics assessment (Northwest Evaluation Association, 2015).

**Table 3.** Growth (and standard deviations) observed in the study relative to the growth norms from the beginning to end of the academic school year on the MAP mathematics assessment for each grade level.

<table>
<thead>
<tr>
<th></th>
<th>Kindergarten</th>
<th>Grade 1</th>
<th>Grade 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Freckle Growth</strong></td>
<td>25.71 (7.51)</td>
<td>25.52 (7.73)</td>
<td>20.68 (8.19)</td>
</tr>
<tr>
<td><strong>Comparison Growth</strong></td>
<td>23.08 (7.17)</td>
<td>22.56 (7.16)</td>
<td>12.79 (5.11)</td>
</tr>
<tr>
<td><strong>MAP Growth Norms</strong></td>
<td>19.1 (7.59)</td>
<td>18.4 (7.45)</td>
<td>15.2 (7.11)</td>
</tr>
</tbody>
</table>

**DATA ANALYSIS**

Student achievement data from the MAP assessment were analyzed by WestEd. The primary analytic strategy for assessing student achievement scores involved fitting Hierarchical Linear Models (HLM). This analysis was chosen because it accounts for the nested structure of the data (i.e., students nested within teachers; Raudenbush & Bryk, 2002) (see Appendix A).

Freckle also conducted interviews with a subset of Freckle teachers to gain insight into the perceived value of Freckle (a summary of the teacher interviews is included in Appendix B).

**RESULTS**

**BASELINE EQUIVALENCE**

An analysis was first conducted to determine whether students were statistically equivalent on their beginning of year MAP scores. This analysis indicated that students in the Freckle condition scored higher ($M_{FR} = 161.44$, $SE_{FR} = 4.22$), but not statistically higher, on the pre-MAP assessment than students in the Comparison condition ($M_{Comp} = 156.05$, $SE_{Comp} = 4.77$; $p = .41$).

To measure the relative size of the baseline differences between the groups, an effect size statistic (i.e., *hedges g*) was calculated. The effect size statistic is a measure of the size of the differences between the groups in standard deviations. The effect size difference between the Freckle and Comparison groups’ baseline scores was .31, which is considered substantive (What Works Clearinghouse, 2014). The difference in beginning-of-year performance is large enough that the two groups should be considered non-equivalent (What Works Clearinghouse, 2014).
**POST-Score Analysis**

To examine student achievement for students whose teachers used Freckle relative to students who did not, we used an HLM model that attempted to control for students’ grade level and their beginning of year scores (see Appendix A).

This analysis indicated that students in the Freckle group exhibit an end of year MAP score that is an estimated 4.64 points higher than students in the Comparison group. This difference is considered significant. The effect size for this difference is .29 (see Figure 2 for means of each groups’ end of year MAP scores).

![Figure 3](image-url)  
**Figure 3.** Covariate-adjusted means for each Group’s end of year MAP scores. Error bars represent standard errors of the means.

To interpret the effect size estimate of .29 in a more meaningful way, we converted this statistic using properties of the normal distribution. This analysis suggested that if students in the Comparison group were at the 50th percentile of a normed sample, the Freckle group students would be placed at the 59th-60th percentile of a normed sample (i.e., approximately 9.86 percentage points higher than the comparison students). We also computed an effect size for each grade level and compared this effect size to the typical expected MAP growth in mathematics (Northwest Evaluation Association, 2015). This analysis suggested that the Freckle group students are approximately two months ahead of Comparison group students in mathematics achievement (see Appendix A).

---

2 Results were similar when excluding the Grade 2 students.
**DISCUSSION**

Students in the Freckle group exhibited higher end of year achievement scores than students in the Comparison group. This effect held when controlling for beginning of year scores and grade level. Thus, findings from this study are at least consistent with the hypothesis that using Freckle in the classroom improves student achievement outcomes.

**STUDY LIMITATIONS**

Because this study relied on data collected post-hoc from a naturalistic implementation, there are many issues that may limit the confidence in conclusions. First, the data set provided by the school was limited in that it contained little background information about participating students and teachers, including reasons why teachers chose to use Freckle at the school. This information would have been useful for inferring the generalizability of findings, as well as for exploring factors that may moderate the effectiveness of Freckle. Second, because this study did not utilize random assignment, it is possible that variables other than Freckle use - such as more effective teachers opting to use Freckle - contributed the increase in student scores. Third, students in the Freckle group had baseline MAP scores that were substantively higher than the Comparison group and therefore cannot be considered equivalent prior to the intervention. Finally, since the study was conducted on a relatively homogeneous and narrow population (participants within a single school), the effects may be larger than would be observed if the study were conducted on a diverse population (e.g., multiple schools in different geographic regions) (Lipsey et al. 2012).

**CONCLUSIONS**

This study is exploratory in nature and represents an initial step in understanding the effectiveness of Freckle on student achievement outcomes. These findings support the hypothesis that Freckle improves students’ mathematics outcomes. However, future studies can utilize random assignment to groups and collect baseline and implementation information to more conclusively determine Freckle’s impact on student outcomes.
REFERENCES


Appendix A.

Statistical and methodological details of the methods and analyses conducted.

**DATA ANALYSIS**

Student achievement data from the MAP assessment were analyzed by WestEd. The primary analytic strategy for assessing student achievement scores involved fitting Hierarchical Linear Models (HLM) with a random effect term for teachers (Raudenbush & Bryk 2002). This analysis was chosen as it accounts for the nested structure of the data (i.e., students nested within teachers). Models were fit using Restricted Maximum Likelihood Estimation (REML). Fixed effects in the model include treatment group, students’ baseline achievement, and grade level. We used R statistical software (R Core Team, 2015) version 3.1.2 and the lme4 package (Bates et al. 2015) version 1.1-7 to conduct the HLM analyses on student achievement.

**QUANTITATIVE ANALYSES**

WestEd analyzed student MAP scores using Hierarchical Linear Modeling (Raudenbush & Bryk 2002). Preliminary model testing indicated that HLM is a suitable choice, as the teacher intraclass correlation was sizeable (ICC = .69) in the unconditional model.

**Baseline Equivalence.** An analysis was first conducted to determine whether students were statistically equivalent at baseline on the pre MAP score. To examine whether students exhibited similar pre MAP scores, student pre MAP Scores were regressed onto the grouping variable (Freckle or Comparison) while including a random effect term for teachers. This analysis indicated that students in the Freckle condition scored higher ($M_{FR} = 161.44, SE_{FR} = 4.22$), but not statistically significantly higher, on the pre-MAP assessment than students in the Comparison condition ($M_{Comp} = 156.05, SE_{Comp} = 4.77; p = .41$). The effect size for this difference was $hedges g = .31$, which is considered substantive. The What Works Clearinghouse (2013) standards considers this level of baseline difference to be non-equivalent.

**Post-Score Analysis.** To examine student achievement for students whose teachers used Freckle relative to students who did not, we used an HLM model which included fixed effects terms for grade and beginning of year scores, as well as a random a effect term to account for the nesting of students within teachers. The HLM model that was used is presented below:

$$PostScore_{ij} = \beta_{00} + \beta_{01}Freckle + \beta_{10}PreScore_{ij} + \beta_{02}FirstGrade_{ij} + \beta_{03}SecondGrade_{ij} + \xi_{0j} + \epsilon_{ij}$$

where $PostScore_{ij}$ is the end of year mathematics score on the MAP assessment for the $i$-th student of the $j$-th teacher, $\beta_{00}$ is the grand mean of scores, $\beta_{10}PreScore_{ij}$ represents students’ beginning of year MAP score, and $\beta_{02}FirstGrade_{ij}$ and $\beta_{03}SecondGrade_{ij}$ represent dummy variables for 1st and 2nd grade classrooms, respectively (with Kindergarten serving as the baseline). $\xi_{0j}$ is a random
Freckle and Student Mathematics Achievement  

The effect term for teachers and and $\epsilon_{ij}$ is a random error term. The difference between scores of students of teachers using Freckle vs. students of teachers in the Comparison group is captured by $\beta_{01,Freckle}$. 

A summary of the random and fixed effects of this model are presented in Tables 4 and 5, respectively.

**Table 4. Random Effects Table.**

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher</td>
<td>5.62</td>
<td>2.37</td>
</tr>
<tr>
<td>Residual</td>
<td>46.08</td>
<td>6.79</td>
</tr>
</tbody>
</table>

**Table 5. Fixed Effects Table.**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>St. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>54.62</td>
<td>4.89</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>PreScore</td>
<td>0.77</td>
<td>0.03</td>
<td>&lt; .001***</td>
</tr>
<tr>
<td>FirstGrade</td>
<td>5.05</td>
<td>1.57</td>
<td>.003**</td>
</tr>
<tr>
<td>SecondGrade</td>
<td>0.66</td>
<td>1.86</td>
<td>0.72</td>
</tr>
<tr>
<td>Freckle</td>
<td>4.64</td>
<td>1.19</td>
<td>&lt; .001***</td>
</tr>
</tbody>
</table>

*Note. **p < .01, ***p < .001*

If the hypothesis about Freckle improving students' mathematics scores is correct, the output in Table 4 suggests that the students in the Freckle group would have grown an estimated 4.64 points higher than students in the Comparison group had the groups been fully equivalent prior to treatment. This difference is significant at the $\alpha < .001$ significance level. The effect size for this difference is *hedge's g* = .29. The covariate-adjusted means for the Freckle and Comparison groups are 184.52 and 179.88, respectively, and the unadjusted standard deviations for these groups are 16.12 and 16.45, respectively.

Because Grade 2 had substantially different sample sizes in the Freckle group relative to the Control group, we reran the above model excluding Grade 2 students. The model coefficient for Freckle term using this subsetted data was still positive, and significant ($B = 3.93, p = .007$).
To interpret the effect size estimate of .29 in a more meaningful way, we converted this statistic using properties of the normal distribution. In the normal distribution, a 1 standard deviation increase corresponds to a 34 point percentile increase. Thus, in a normed sample, an effect size of .29 corresponds to a 9.86 point percentile increase (i.e., $34 \times .29 = 9.86$). Therefore, the results suggest that if students in the Comparison group were at the 50th percentile of the normed sample, the Freckle group students would be placed at the 59th - 60th percentile of the normed sample (i.e., $50 + 9.86 = 59.86$).

We also computed an effect size for each grade level and compared this effect size to the typical expected MAP growth in mathematics. To this end, we reran the model presented above separately for each grade level (i.e., we excluded the grade-level terms from each model). The effect sizes for the Freckle variable from these models are .29, .38, and .34 for Kindergarten, Grade 1, and Grade 2, respectively. Based on the beginning and end of year student status norms on the MAP mathematics assessment (Northwest Evaluation Association, 2015), the estimated beginning to end of year effect sizes for MAP mathematics are 1.33, 1.39, 1.14 for each of these grades. Comparing the effect sizes observed in the present study to the effect sizes of the average growth for the MAP assessment suggests that the observed effect sizes are noteworthy, corresponding to 22%, 27%, and 30% of the expected gain over the course of a school-year in Kindergarten, Grade 1, and Grade 2, respectively (e.g., for Kindergarten, $.29/1.33 = .22$, etc.). Multiplying the average school-year gain for all grade-levels (.26) by 9 months (the typical number of months in a school year), we estimate that the Freckle group students are approximately two months ahead of Comparison group students in mathematics achievement.
Appendix B.

Description of teacher interviews conducted by Freckle.

**Teacher Interviews**

Teachers’ perceived value of the program was defined as the extent to which Freckle could be used effectively and efficiently by classroom teachers. To gain an understanding of the perceived value of Freckle, Freckle conducted interviews with three teachers who used Freckle about their experience with the program. The interviews were conducted over the phone in the fall of 2015, and consisted of questions about how Freckle usage impacted student engagement and outcomes in math. The teachers interviewed had used Freckle in the 2014-2015 school year and were willing to discuss their experiences with the program. Questions included how they felt Freckle impacted growth in their classroom, students’ engagement in math, and their effectiveness as teachers. Interviews lasted approximately 30 minutes each.

In general, Freckle was used in math centers with small groups of students for about 30 minutes a day, a few times a week. All the teachers interviewed said they would recommend Freckle, citing its individualized practice as the main benefit of the program. The interviewed teachers reported that they saw tremendous growth from using Freckle. For instance, one teacher said, “at the beginning of the year last year, I had kids who were significantly below grade level. By the end of the year, out of the kids I had, only 1 or 2 were not on grade level or above”.

Teachers also appreciated the way Freckle deepened students’ understanding of math concepts. One teacher remarked that “it built their confidence, because the problems were huge, and once they could break it down into its parts, they got it. They used strategies to solve the problem. I believe it exposed them to thinking outside the box”. Another teacher reported that students enjoyed it so much that they wanted to use Freckle at home.

Teachers reported that Freckle allowed them to better meet the needs of every child in their classroom, giving them the ability to focus on small groups of students at a time. One teacher stated that Freckle “allows me to see what I don’t need to waste time on. I’m able to better manage students’ instruction and manage my time better”.

Overall, the interviewed teachers viewed Freckle as an effective tool, with one saying “once I found out about it, I basically used it all the time”.

Interview responses suggest that Freckle can be a valuable classroom tool. Though only a small subset of teachers that used Freckle participated in interviews, these teachers suggested that Freckle was fluidly integrated into their classroom instruction, and that they were able to use class time more effectively to support student mathematics learning.