



Evaluation of the TrueSuccess® Program

Effects of Participation in the
TrueSuccess® Program on Key
Educational Outcomes

November 2021

Prepared for:

TrueSuccess® Inc.



RE Analytics

William Medendorp Ph.D., Principal Investigator

Walter DeBoer, Ed.D., Co-Investigator

Kevin Anderson, Ph.D., Co-Investigator



Contents

Acknowledgements	1
Executive Summary	2
Program Description	2
TrueSuccess® provides research-based educational tools, training and support that are designed to equip children with social and emotional skills. At the core of TrueSuccess are its No Prep Lessons, which include engaging and inspiring virtue videos and life stories of positive role models.....	2
Population Served	2
Findings	2
<i>Fall to spring MAP growth scores of TrueSuccess participants vs. comparison group.</i>	<i>2</i>
<i>Suspension rates of TrueSuccess participants vs. comparison group.</i>	<i>2</i>
<i>Differences in academic and behavioral outcomes of TrueSuccess participants depending on school, grade level, and demographics.</i>	<i>2</i>
<i>Differences in academic and behavioral outcomes depending on level of implementation of TrueSuccess.</i>	<i>2</i>
<i>Differences in academic and behavioral outcomes for participants who participate in TrueSuccess programming for two consecutive years compared to those who participate for only one year.</i>	<i>2</i>
Conclusions.....	2
Program Description.....	3
Purpose of this Study.....	3
Research Questions	5
Program Implementation	5
Population	6
Research Methodology.....	7
Mahalanobis Distance Matching (MDM)	7
<i>Approach #1: Single Fidelity Threshold</i>	<i>7</i>
<i>Approach #2: Multiple Fidelity Thresholds.....</i>	<i>7</i>
<i>Approach #3: Multiyear, Single Fidelity Threshold</i>	<i>7</i>
Matched Comparisons using Linear Mixed Effects Regression (LMER)	8
Findings.....	9
<i>Do students who participate in TrueSuccess programming achieve greater fall to spring growth in reading and math than comparison group students who do not participate in TrueSuccess?.....</i>	<i>9</i>
<i>Do students who participate in TrueSuccess programming have lower rates of suspensions than comparison group students who do not participate in TrueSuccess?</i>	<i>9</i>

Are there statistically significant differences in outcomes for participants (achievement growth and suspensions) depending on school, grade level, or demographics?10

Are there statistically significant differences in outcomes based on the level of implementation of TrueSuccess?.....11

Are there statistically significant differences in outcomes for participants who participate in TrueSuccess programming for two consecutive years compared to those who participate for only one year?.....12

Conclusions 14

Appendix A: Extended Demographic Information..... 15

Appendix B: Mahalanobis Distance Matching and Model Creation Methods 16

 Mahalanobis Distance Matching (MDM) 16

 Approach #1: Single Fidelity Threshold 16

 Approach #2: Multiple Fidelity Thresholds..... 16

 Approach #3: Multiyear, Single Fidelity Threshold..... 16

 Model Creation..... 17

 Model and Predictor Assessment 17

Appendix C: Matched Comparison Group Results..... 19

 Approach #1: Single Fidelity Threshold..... 19

 Approach #2: Multiple Fidelity Thresholds 19

 Approach #3: Multiyear, Single Fidelity Threshold 21

 Fidelity as a Factor..... 22

Appendix D: TrueSuccess School Participation 23

Appendix E: TrueSuccess® Promotional Overview 24

 MISSION 24

 HISTORY OF SUCCESS..... 24

 OFFERING 24

Appendix F: Bibliography 26

Acknowledgements

Key Evaluation Contacts

TrueSuccess®

Bill Heneveld, TrueSuccess® Inc. Founder

Dr. Art Garner, President

Grand Rapids Public Schools

Ronald Gorman, Ph.D., Deputy Superintendent

Mel Atkins II, Executive Director of Community & Student Affairs

Alexandra Watson, Grant and Research Specialist

Katherine Pitsch, Educational Research Analyst

Executive Summary

Program Description

TrueSuccess® provides research-based educational tools, training and support that are designed to equip children with social and emotional skills. At the core of TrueSuccess are its No Prep Lessons, which include engaging and inspiring virtue videos and life stories of positive role models.

Population Served

This study focuses on TrueSuccess' implementation in seven schools within a large public district in the Midwest over the course of 2 school years. The population sample includes a total of 8,320 students in grades K-5. Of these students, 1,982 students received at least one year of instruction from teachers trained and schools supported by TrueSuccess.

Findings

Fall to spring MAP growth scores of TrueSuccess participants vs. comparison group.

Students in TrueSuccess demonstrate significantly higher annual growth in math and reading scores compared to their peers outside the program.

Suspension rates of TrueSuccess participants vs. comparison group.

Students in TrueSuccess receive 20% fewer suspensions than their peers outside the program.

Differences in academic and behavioral outcomes of TrueSuccess participants depending on school, grade level, and demographics.

Improvements attributable to TrueSuccess remain independent of other factors such as gender, grade, ethnicity/race, etc.

Differences in academic and behavioral outcomes depending on level of implementation of TrueSuccess.

Increased TrueSuccess exposure reveals little additional impact on math and reading improvements. Suspensions decrease with increasing TrueSuccess exposure.

Differences in academic and behavioral outcomes for participants who participate in TrueSuccess programming for two consecutive years compared to those who participate for only one year.

Multiyear enrollment yields mixed results. This is due to not having access to raw, unstandardized scores for spring semesters, making mathematical comparison misleading.

Conclusions

- TrueSuccess clearly has a positive, statistically significant impact on both academic (reading and math) and behavioral (suspensions) metrics. This is true regardless of other factors (e.g., gender, grade, race/ethnicity).
- TrueSuccess' positive impact on academic performance (reading and math) remains significant regardless of implementation level.
- TrueSuccess' significantly positive impact on behavior (reducing suspension by 20%) requires implementing 12 or more TrueSuccess lessons per month.

Program Description

TrueSuccess® was developed by founder Bill Heneveld to “help reverse the culture’s character crisis by inspiring kids with character to unleash their potential in school, work and life.”¹ TrueSuccess provides research-based, grade level appropriate, educational tools, training and support that are designed to equip children with social and emotional skills essential to unleashing their potential. TrueSuccess includes: 1) a curriculum (both online and printed, for grades K-12), 2) initial educator training, 3) educator coaching (school year-long support), 4) administrator support including analytics of tool use and impact, 5) SEL assessment surveys for students and teachers, and 6) adaptation interventions to address assessed deficiencies. The curriculum is organized around positive behavior skills, is fun and educational. At the core of TrueSuccess are its No-Prep Lessons, which include engaging, inspiring and impactful virtue videos and life stories of positive role models who exemplify positive behavior skills that encourage overcoming life’s challenges. The curriculum is refreshed every school year with new content and lessons. The curriculum is flexible and uses have included in-classroom, after-school, mentoring, parenting, community centers and camps. TrueSuccess has been distributed to over 230,000 children/youth in grades K-12.

TrueSuccess is designed to equip positive behavior skills including: Respect, Responsibility, Gratitude, Self-Control, Grit, Courage, Encouragement, Compassion, and Integrity. These skills align closely with five social and emotional learning (SEL) competencies identified by the Collaborative for Academic, Social, and Emotional Learning (CASEL). Also, TrueSuccess’ pre, mid and post assessment surveys for students and teachers assess CASEL’s five SEL competencies. CASEL’s stated mission is to “help make evidence-based social and emotional learning an integral part of education from pre-school through high school.”² CASEL refers to these competencies as the “CASEL 5”: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. CASEL defines SEL as “the process through which all young people and adults acquire and apply the knowledge, skills, and attitudes to develop healthy identities, manage emotions and achieve personal and collective goals, feel and show empathy for others, establish and maintain supportive relationships, and make responsible and caring decisions.”³

Purpose of this Study

This research study seeks to determine whether participation of students in the TrueSuccess program during the 2017-18 and 2018-19 school years has a statistically significant impact on fall to spring reading and math growth and number of suspensions when compared to equivalent non-participants. TrueSuccess was implemented in seven elementary schools in the Grand Rapids Public School (GRPS) district in grades K-5. These

¹ <https://truesuccesstools.org/mission-history/>

² <https://casel.org/about-us/our-mission-work/>

³ <https://casel.org/fundamentals-of-sel/>

two school years were chosen because 1) they precede the COVID-19 pandemic and attendant interruptions in educational delivery, 2) fidelity logs were consistently maintained by all teachers implementing TrueSuccess for both years, and 3) full data sets of all data elements were readily available from archived records for both years.

This study has the potential to benefit the school district by confirming that a program/intervention currently in use meets rigorous evidence-based standards with benefits for at least one of two key district priorities: student academic achievement and classroom behavior. If confirmed, principals and staff at other district schools might wish to invest in the TrueSuccess program.

The study also has the potential to benefit those school districts, after-school programs, and other institutions that focus their programming on the social-emotional learning (SEL) growth of children and youth.

CASEL has established a large body of research that shows that students benefit academically, socially, and behaviorally when they have well developed social-emotional learning competencies.⁴ CASEL identifies the following criteria to be designated as evidence-based CASEL “SElect” programs:

- Be well-designed classroom-based programs that systematically promote students’ social and emotional competence, provide opportunities for practice, and offer multi-year programming.
- Deliver high-quality training and other implementation supports, including initial training and ongoing support to ensure sound implementation.
- Be evidence-based with at least one carefully conducted evaluation that documents positive impacts on student behavior and academic performance.

TrueSuccess meets the first two standards. Conducting a rigorous evaluation that meets the third evidentiary standard required access by Research & Evaluation Associates (REA) to de-identified student data from the 2017-18 and 2018-19 school years from the Grand Rapids Public School district. REA followed the district’s data-sharing protocols and was approved to receive student-level data elements for each of the two school years. These data elements included fall and spring MAP RIT (Rasch Unit) scores in reading and math; MAP fall to spring CGI (Conditional Growth Index) growth scores; school attended on fall count day; teacher name; school attended at year end; grade level, ethnicity/race, sex, and qualification for free/reduced lunch (y/n); number of suspensions; number of days enrolled; and number of days attended. Students who participated in TrueSuccess were identified by school, teacher name, and grade level by matching against fidelity logs that

⁴ <https://casel.org/research/>

track teacher use of online lessons for both years. Propensity matching was used to establish a statistically equivalent group by approximating random assignment.

Research Questions

The analysis seeks to address the following research questions:

1. Do students who participate in TrueSuccess programming achieve greater fall to spring growth in reading and math than comparison group students who do not participate in TrueSuccess?
2. Do students who participate in TrueSuccess programming have lower rates of suspensions than comparison group students who do not participate in TrueSuccess?
3. Are there statistically significant differences in outcomes for participants (achievement growth and suspensions) depending on school, grade level, or demographics?
4. Are there statistically significant differences in outcomes based on the level of implementation of TrueSuccess?
5. Are there statistically significant differences in outcomes for participants who participate in TrueSuccess programming for two consecutive years compared to those who participate for only one year?

Program Implementation

The decisions to purchase and implement TrueSuccess in GRPS were made by individual school principals. Over a period of time, founder Bill Heneveld approached the principals at each of the seven schools in this study, promoting the benefits of TrueSuccess, and requesting the opportunity to present on a more formal basis. In some instances, these follow-up presentations were with the principal and in others they were with the principal, school staff, and classroom teachers. The funds to purchase TrueSuccess came from a grant secured by TrueSuccess and discretionary instructional accounts that principals manage. Principals of schools designated as Title 1 or Section 31a could use grant funds and supplemental instructional funds to purchase TrueSuccess. Initial teacher training plus school year-long support of teachers, principals, and school staff was provided by TrueSuccess in all seven schools. Having invested in the program, principals urged their teachers to use TrueSuccess at the levels recommended; however, implementation was at teachers' discretion. Therefore, fidelity to the TrueSuccess model varied by school and classroom. Levels of use are tracked automatically as teachers log in and access on-line lessons. TrueSuccess then produces Teacher-Use Analytics reports to document use.

Population

The TrueSuccess program was implemented in seven schools within Grand Rapids Public Schools (GRPS), a large public school system in the Midwest, in grades K-5 over the course of two school years (2017-18 & 2018-19). Data were collected over this timeframe for a total of 8,320 district students in grades K-5. Of these students, 1,982 had received at least one year of instruction from teachers trained and schools supported by the TrueSuccess program.

Table 1 shows student demographics regardless of TrueSuccess enrollment. The 17-18 school year file includes a total of 6,644 students, while the 18-19 school year file includes 6,358 students. A total of 1,008 students participated in both school years. Table 1 indicates roughly equal distribution among grade levels. Hispanics represent the largest race/ethnicity, followed by African American/Black. Most students receive free or reduced lunch.

Table 1: Unmatched Demographics of GRPS Sample

	Unmatched		Matched	
	N	%	N	%
Total	10401	100%	1048	100%
Grade				
K	1758	16.9%	144	13.7%
1	1705	16.4%	151	14.4%
2	1698	16.3%	146	13.9%
3	1715	16.5%	204	19.5%
4	1787	17.2%	223	21.3%
5	1738	16.7%	180	17.2%
Gender				
M	5326	51.2%	551	52.6%
F	5075	48.8%	497	47.4%
Race/Ethnicity				
Hispanic	3451	33.2%	681	65.0%
Black	3740	36.0%	212	20.2%
White	2294	22.1%	85	8.1%
2 or More	916	8.8%	70	6.7%
FRL				
No	3780	36.3%	275	26.2%
Yes	6621	63.7%	773	73.8%

Note. FRL = Free/Reduced Lunch, N = # of Students.
 Matched group includes students from unmatched control group that matched TrueSuccess students
 Low case count variables removed

Figure 1: Race and Ethnicity of Matched Students

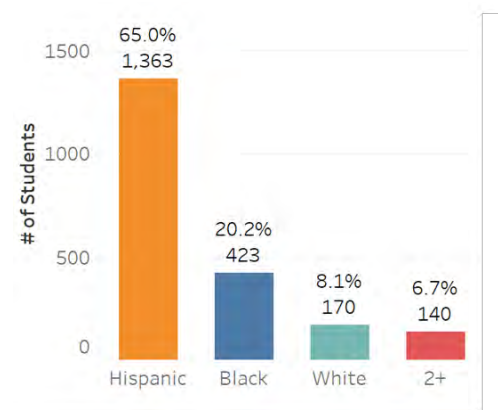
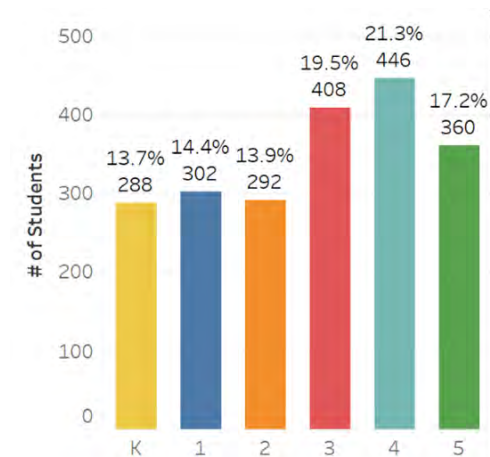


Figure 2: Grade level of Matched students



Research Methodology

Mahalanobis Distance Matching (MDM)

A quasi-experimental matched comparison group design provides meaningful differences between the TrueSuccess and nonparticipant groups. With an accurate matching protocol, comparison group analysis tests for significant differences between TrueSuccess participants and equivalent nonparticipants, which can be assumed to be due to the TrueSuccess intervention. Using Mahalanobis distance (Mahalanobis Distance Matching, MDM), students from TrueSuccess classrooms were matched with an equal number of non-TrueSuccess peers (hereto referred to as “the control group”) controlling for demographics, baseline RIT scores, and specific characteristics.

To answer the research questions, the analysis presents 3 separate matching cohorts based on the number of no-prep lessons taught. Teacher Use Analytics provide counts of no-prep lessons delivered in October and May of the 17-18 and 18-19 school year. These months were chosen as a proxy for implementation level due to the limited number of breaks and days off during those months. Teachers are assigned a “fidelity measure” based on average number of no-prep lessons implemented during each of these months. Students receive the fidelity measure assigned to their Lead Teacher.

The following cohorts were matched:

Approach #1: Single Fidelity Threshold

This approach consists of students in classrooms that received an average of 12 or more no-prep lessons per month. TrueSuccess staff indicate a threshold of 12 lessons per month ensures sufficient exposure to the TrueSuccess program. Table 2 shows a total of 1,048 students meet this criterion and were matched to 1,048 students from the control group. Select demographics for these 2,096 students are shown in Figures 1 & 2, while complete demographics for the matched groups are given in Appendix A.

Approach #2: Multiple Fidelity Thresholds

To assess how the *level* of implementation affects performance (Research Question #5), students were matched at varying fidelity thresholds ranging from 0 to more than 18 lessons per month. Table 2 reveals the total number of students included in the analysis at each fidelity threshold.

Approach #3: Multiyear, Single Fidelity Threshold

To analyze the impact of participation in TrueSuccess for both the 17-18 and 18-19 school year compared to those with only a single year (Research Question #4), students represented in both school years were matched with control group students. The analysis considered 4

Table 2: Sample at each Fidelity Threshold

Fidelity	N
0	1968
3	1499
6	1370
9	1182
12	1048
15	878
18	709

separate groups: no TrueSuccess in either year, TrueSuccess in 17-18, TrueSuccess in 18-19, and TrueSuccess in both years. All students in the TrueSuccess groups received an average of at least 12 no-prep lessons of TrueSuccess. The sample consists of 266 students per group for a total of 1,064 students.

For more details on the matching process, see Appendix B.

Matched Comparisons using Linear Mixed Effects Regression (LMER)

The analysis utilizes a linear mixed effects model to compare the matched groups and to assess differences between TrueSuccess and non-TrueSuccess students with 3 primary outcomes: math Conditional Growth Index (CGI) scores, reading CGI scores, and number of suspensions. The model controls for student gender, ethnicity/race, grade, free/reduced lunch as fixed effects and student ID and school as random effects.

The analysis implements the mixed effects regression on all match approaches, as well as a unique analysis using fidelity measures as a predictor. Details on model specification can be found in Appendix C.

Findings

Do students who participate in TrueSuccess programming achieve greater fall to spring growth in reading and math than comparison group students who do not participate in TrueSuccess?

TrueSuccess students with a minimum fidelity measure of 12 earn significantly higher math and reading growth scores than their non-TrueSuccess peers. Table 3 shows TrueSuccess participation significantly differs compared to matched nonparticipants, when controlling for student demographics and characteristics. Figure 3 indicates TrueSuccess math CGI scores average 0.17 points compared to -0.01 among nonparticipants, while reading scores average -0.05 points compared to -0.15 among nonparticipants. This indicates TrueSuccess participants show improved academic performance.

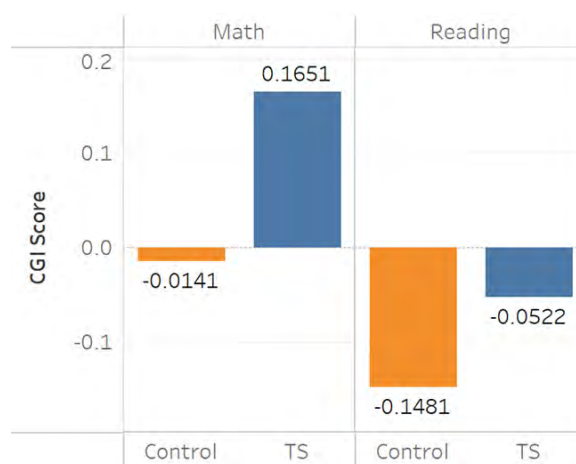
Note: ANOVA results provide post-hoc testing between groups. A significant value indicates significant differences between groups were found (i.e. TrueSuccess is significantly different from nonparticipants).

Table 3: ANOVA results from LMER reveals Significantly higher math and reading CGI scores among TrueSuccess participants, N = 1048 per group

	<i>MSE</i>	<i>F</i>	<i>p</i>
Math			
TS	7.58	5.65	*
Grade	30.62	22.82	***
FRL	5.18	3.86	*
Gender	4.27	3.18	-
Attendance	32.39	24.13	***
Reading			
TS	5.09	5.27	*
Grade	4.79	4.96	***
FRL	0.49	0.50	-
Gender	2.10	2.18	-
Attendance	8.05	8.34	**

Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$
 MSE = mean square error, TS = TrueSuccess
 FRL = free/reduced lunch

Figure 3: TrueSuccess participants show higher math and reading CGI scores compared to matched nonparticipants, N = 1048 per group



Do students who participate in TrueSuccess programming have lower rates of suspensions than comparison group students who do not participate in TrueSuccess?

Students from TrueSuccess classrooms receive significantly fewer suspensions than matched nonparticipants. Table 4 shows TrueSuccess participants report significantly reduced number of suspensions compared to matched nonparticipants ($\beta = -0.50$, $p < 0.05$). Over the course of two years TrueSuccess participants receive a total of 208 suspensions

compared to 253 suspensions in the Control group, approximately 20% fewer suspensions. See Figure 4

Figure 4: TS Participants show significantly fewer suspensions

Table 4: GLMM Results indicate TrueSuccess participants significantly reduce suspensions, N = 1048 per group

	<u>β</u>	<u>Error</u>	<u>z</u>	<u>p</u>
TS	-0.50	0.25	-1.97	*
Gender	1.24	0.35	3.60	***
Race/Ethnicity				
Hispanic	-1.44	0.35	-4.12	***
White	-0.97	0.59	-1.64	-
2+	-0.15	0.56	-0.26	-



Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$

Error = standard error, TS = TrueSuccess

Reference group for ethnicity is “Black”

Reference group for gender is “Female”

Are there statistically significant differences in outcomes for participants (achievement growth and suspensions) depending on school, grade level, or demographics?

Participant outcomes show no significant variation by grade, gender, race/ethnicity or FRL enrollment. While the data show significant differences exist between different levels of the previously stated predictors, TrueSuccess participation shows equal impact across all levels of these predictors.

Note: Inclusion of School as a random effect, nesting TrueSuccess participation within school designation, makes school comparisons redundant and meaningless. For more details about how students were distributed between schools, see Appendix D.

Are there statistically significant differences in outcomes based on the level of implementation of TrueSuccess?

The number of suspensions experienced by participants significantly decrease as more TrueSuccess lessons are implemented, while average growth in math and reading scores shows no dependence on teacher fidelity. Table 5 reveals no consistent patterns in math and reading CGI scores based on different levels of fidelity. Participant suspensions, by contrast, significantly decrease with increasing TrueSuccess fidelity.

Note: The lack of clear trend may be due to the use of a single month as a proxy for implementation throughout the semester. Review of monthly variation indicates the measures used are a reasonable proxy yearly implementation; however, a more detailed analysis of Teacher Use Analytics might define these relationships more clearly.

Table 5: Suspensions show a clear decline with increased participation, while Math and Reading CGI scores do not reveal a consistent participation-dependent pattern

	β	Error	df	t/z	p
Math					
0	0.18	0.07	1610	2.78	**
3	0.21	0.08	1032	2.76	**
6	0.16	0.08	926	1.93	-
9	0.22	0.09	504	2.42	*
12	0.22	0.09	454	2.38	*
15	0.19	0.10	386	1.83	-
18	0.16	0.12	146	1.32	-
Reading					
0	0.16	0.06	1113	2.81	**
3	0.18	0.07	673	2.64	**
6	0.14	0.07	532	1.89	-
9	0.16	0.08	385	1.98	*
12	0.20	0.09	367	2.29	*
15	0.28	0.10	350	2.85	**
18	0.25	0.12	162	2.10	*
Suspensions					
0	-0.02	0.14	-	-0.12	-
3	-0.16	0.17	-	-0.93	-
6	-0.12	0.18	-	-0.63	-
9	-0.23	0.20	-	-1.15	-
12	-0.50	0.25	-	-1.97	*
15	-0.83	0.30	-	-2.80	**
18	-1.02	0.33	-	-3.12	**

Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$

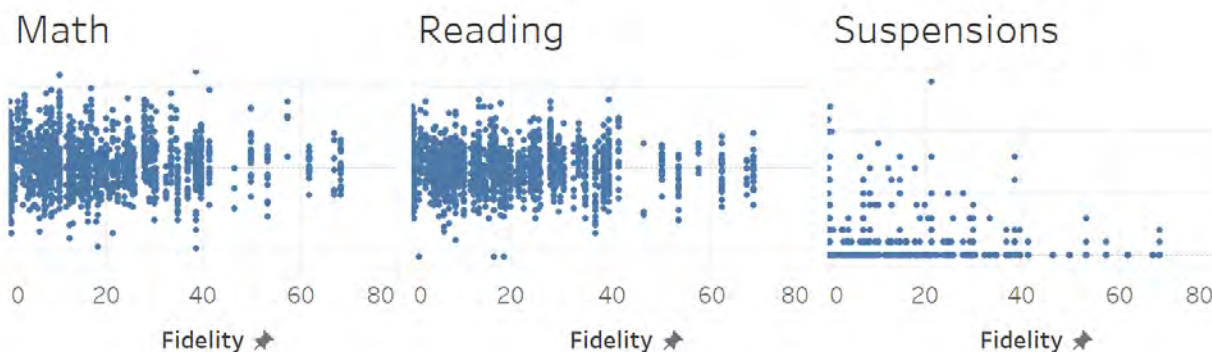
Error = standard error, df = degrees of freedom

Math/Reading LMER models use t

Suspension GLMM model uses z (no df)

Figure 5 shows no significant trends with increasing fidelity among math and reading CGI scores. Suspensions, by contrast, significantly decrease with increasing fidelity. A fidelity measure of 12 appears to be the point where TrueSuccess students begin to show reduced suspensions compared to matched peers.

Figure 5: Increased participation in TrueSuccess is associated with significantly reduced suspensions, while reading and math CGI scores show no trends based on fidelity



Are there statistically significant differences in outcomes for participants who participate in TrueSuccess programming for two consecutive years compared to those who participate for only one year?

Note: This analysis utilizes CGI scores that are calculated based on expected (normative) growth for a single year. They cannot readily be transformed for use in a multiyear growth analysis without the original RIT scores. Due to this, the above findings do not identify a clear pattern, nor do they seem to match the results of the single-year analyses. The variability of the resulting means, as well as the lower sample size of the two-year participation, suggests there simply was not enough consistent data to draw any precise conclusions.

Participants enrolled in TrueSuccess for 2 consecutive years show minor differences in outcomes compared to participants enrolled for one year. Table 6 shows comparisons between the 17-18 TrueSuccess enrolled students, 18-19 TrueSuccess enrolled students, students enrolled in TrueSuccess for both school years, and matched nonparticipants. Students from the two-year TrueSuccess groups perform significantly better on math CGI scores than their peers from the control group and the 17-18 TrueSuccess group at an average point difference of 0.21. Participants from the 2-year TrueSuccess groups show no significant improvement from single year TrueSuccess groups or the control group on reading CGI scores or on suspensions.

Table 6: Post-hoc comparisons show significantly improved math CGI scores for multi-year enrolled TrueSuccess participants

	<u>MSE</u>	<u>df₁</u>	<u>df₂</u>	<u>F</u>	<u>p</u>
Math					
TSMulti	3.95	3	181	3.03	*
Grade	13.26	5	483	10.18	***
FRL	0.14	1	481	0.11	-
Gender	0.92	1	481	0.71	-
Attendance	3.39	1	483	2.60	-
Reading					
TSMulti	0.53	3	173	0.33	-
Grade	1.74	5	484	1.10	-
FRL	0.70	1	482	0.45	-
Gender	0.72	1	482	0.46	-
Attendance	2.08	1	484	1.32	-

Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$

MSE = mean square error, df_1 = numerator degrees of freedom

df_2 = denominator degrees of freedom,

TSMulti = TrueSuccess, Multiyear, FRL = free/reduced lunch

Conclusions

Based on the findings summarized in the preceding section, REA offers the following conclusions for consideration.

- TrueSuccess clearly has a positive, statistically significant impact on both academic (reading and math) and behavioral (suspensions) metrics. This is true regardless of other factors (e.g., gender, grade, race/ethnicity).
- TrueSuccess' positive impact on academic performance (reading and math) remains significant regardless of implementation level.
- TrueSuccess' significantly positive impact on behavior (reducing suspension by 20%) requires implementing 12 or more TrueSuccess lessons per month.

Appendix A: Extended Demographic Information

Table 7: Demographics by Matched and Unmatched cases

	<u>Ctrl, Unmatched</u>		<u>TS, Fidelity 0</u>		<u>Ctrl, Matched</u>		<u>TS, Fidelity 12</u>	
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
Total	10401	100%	1968	100%	1048	100%	1048	100%
Grade								
K	1758	16.9%	407	20.7%	144	13.7%	144	13.7%
1	1705	16.4%	347	17.6%	151	14.4%	151	14.4%
2	1698	16.3%	301	15.3%	146	13.9%	146	13.9%
3	1715	16.5%	316	16.1%	204	19.5%	204	19.5%
4	1787	17.2%	281	14.3%	223	21.3%	223	21.3%
5	1738	16.7%	316	16.1%	180	17.2%	180	17.2%
Gender								
M	5326	51.2%	1022	51.9%	551	52.6%	551	52.6%
F	5075	48.8%	946	48.1%	497	47.4%	497	47.4%
Race/Ethnicity								
Hispanic	3451	33.2%	1315	66.8%	681	65.0%	682	65.1%
Black	3740	36.0%	372	18.9%	212	20.2%	211	20.1%
White	2294	22.1%	157	8.0%	85	8.1%	85	8.1%
2 or More	916	8.8%	124	6.3%	70	6.7%	70	6.7%
FRL								
No	3780	36.3%	572	29.1%	275	26.2%	276	26.3%
Yes	6621	63.7%	1396	70.9%	773	73.8%	772	73.7%

Note. FRL = Free/Reduced Lunch, N = # of Cases.

Data for “X” gender and “Asian”, “Native American”, and “Pacific Islander” race/ethnicity categories removed due to low case count

Appendix B: Mahalanobis Distance Matching and Model Creation

Methods

Mahalanobis Distance Matching (MDM)

Prior to any Mahalanobis distance matching (MDM), data was tabulated and low case factor levels were removed to consolidate the data, including the “X” level of Gender and three levels for Race/Ethnicity: “Asian”, “Native American”, and “Pacific Islander”.

All matching procedures were performed in R (R Core Team, 2021) using the “MatchIt” package (Ho, Imai, King, & Stuart, 2011). All matches used optimal pair matching compared to nearest-neighbor matching. In no observed scenarios did nearest-neighbor matching outperform optimal pair matching. The analysis tested comparisons of the use of propensity score matching (PSM) side-by-side with MDM using the same “MatchIt” package; however, resulting matches were at best equivalent to the MDM results.

Three general approaches to PSM were implemented in this study:

Approach #1: Single Fidelity Threshold

The original sample of cases from TrueSuccess classrooms from both the 17-18 and 18-19 school years ($N = 1,698$) was reduced so that only students from classrooms where 12 or more TrueSuccess lessons on average were taught (i.e., fidelity measure = 12; $N = 1,048$). Students participating in multiple years were counted only once for the 17-18 school year, their first exposure to TrueSuccess. These students were matched to an equal number of students that were never in a TrueSuccess classroom using the previously described matching methods.

Approach #2: Multiple Fidelity Thresholds

The process from Approach #1 was repeated from the beginning, but the fidelity measure threshold was set to one of the following values: 0, 3, 6, 9, 15, 18. For each threshold, the students who were in TrueSuccess classrooms that did not meet or exceed the fidelity requirement were not counted as controls but were removed entirely. Students that were in TrueSuccess classrooms for two years were also removed from the matching pool.

Approach #3: Multiyear, Single Fidelity Threshold

The original dataset was reduced to include only students that had any sort of case data for two years. This reduced the data set to 9,280 cases over two years. After this, the data was split into four groups: control ($n = 6,748$), 17-18 TrueSuccess ($n = 532$), 18-19 TrueSuccess ($n = 992$), and two-year TrueSuccess ($n = 1,008$). From here, students were matched according to their data for the 17-18 school year (i.e., half n). Groups were MDM matched again, but since said method only matches two groups at a time, we chose the smallest group (17-18 TrueSuccess) as a reference group against which the other three groups were

matched. The three separate matched datasets were recombined so that there were four equal groups of $n = 532$. From here, the dataset was cut in half to only include 18-19 data ($n = 266$ per group). This final dataset was used for all models moving forward.

Model Creation

Models were created in R using the “lme4” package (Bates, Mächler, Bolker, & Walker, 2015). Math CGI and reading CGI scores were modeled using linear mixed effect regression (LMER), while suspensions were formulated using general liner mixed models (GLMM) based on Poisson regression. Both models feature the separation of fixed effects and random effects. Both have flexible definitions, but one way to frame the difference between them is that fixed effects represent factors where all levels are accounted for while random effects can be viewed as a subset of a larger population. Student ID and school were used as random effects, while TrueSuccess participation (TS), gender, race/ethnicity, free/reduced lunch status (FRL), grade, and attendance are all defined as fixed effects.

From the above list, a variant of forward stepwise selection was implemented to create the final models. Rather than testing all possible additions at all possible stages of selection, the order in which variables were added was established a priori. Initial models contained TS and ID only, then factors were added in the following order: school, race/ethnicity, grade, gender, FRL, attendance. After this, the list continued with interaction effects depending upon which remained in the model. Only interaction effects that included TS were tested. Each addition was followed by a likelihood ratio test (LRT) with the previous model. If $p < .05$, the factor is kept. If $p > .10$, the factor was removed with no further consideration. If $.05 < p < .10$, the factor was added in an earlier step to see if the results changed (earlier additions tend to improve the model more).

This was done for all three measures using matched data from Approach #1 first. Models for math and reading were very similar and thus the same model structure was implemented for both. Suspensions followed a different structure and were modeled separately. Also worth noting is that the models behaved similarly between different matched datasets, which lead to the same general structure from Approach #1 being applied to Approaches #2 and #3. The exact results are given in Appendix C.

Finally, in order to assess how TrueSuccess effects change with different amounts of no-prep lessons taught, LMER/GLMM models were created using unmatched data that included only TrueSuccess students. This model shared structure with the previously defined models and the end results are again defined in Appendix C.

Model and Predictor Assessment

Once the models were formed, models were tested in multiple ways. For the LMER models, two results were calculated: ANOVA results and regression coefficients. The distribution of math and reading scores met assumptions for ANOVA testing, which allowed for simplified results when compared with slopes from the regression for multilevel predictors (e.g.,

grade, race/ethnicity). The ANOVA results were approximated from the LMER model using the “lmerTest” package (Kuznetsova, Brockhoff, & Christensen, 2017). This package uses Satterthwaite’s approximation for degrees of freedom to estimate Type III ANOVA results. The probability values obtained from this method are near equivalent to calculating t for the regression coefficients and can be used interchangeably depending on which is more interpretable within a given context.

For the GLMMs, ANOVA results are not an option due to the use of a non-linear distribution for suspensions. As such, Wald’s test alone was used to evaluate the effect of each predictor in the model.

Appendix C: Matched Comparison Group Results

Approach #1: Single Fidelity Threshold

After matching the groups so that 1,098 TrueSuccess students with fidelity measures of 12 or higher against 1,098 non-TrueSuccess controls, the model using math CGI scores was built using forward stepwise selection until a final LMER model was created:

$$\text{Math/Read CGI} \sim \text{TS} + \text{Grade} + \text{FRL} + \text{Gender} + \text{Attendance} + (1|\text{ID})^5 + (1|\text{School}) \quad (1)$$

Note that Model 1 applies to both math CGI and reading CGI scores separately (i.e., two separate models). Race/Ethnicity and any interaction effects were removed from the model due to insignificant impact on the model. Suspensions being a GLMM with a Poisson distribution yielded the following model:

$$\text{Suspensions} \sim \text{TS} + \text{Gender} + \text{Race/Ethnicity} + (1|\text{ID}) \quad (2)$$

Notably fewer predictors improved the model, although race/ethnicity did significantly improve prediction of suspension rate. It is also the case that (1|School) caused convergence errors in combination with (1|ID) and thus was dropped. As before, interaction effects did not significantly improve the model.

Model results for math/reading are given in Table 8, while ANOVA results are given in Table 9. Regression results for suspensions are given in Table 4 above. Enrollment in TrueSuccess classrooms increases math CGI scores by 0.18 points, increases reading CGI scores by 0.10 points, and decreases suspensions by 20%. When random effects are accounted for by the models, these predictive estimates change to 0.22, 0.20, and 40% respectively.

Approach #2: Multiple Fidelity Thresholds

The above models were fit using data matched using only TrueSuccess cases with fidelity measures greater than or equal to 12. When that threshold is changed and the raw data is rematched and then fit to the same models, the results can be seen in Table 5 above. Notably, there are improvements at most thresholds for both math and reading with no discernible pattern based on increasing fidelity. Suspensions, on the other hand, clearly shows a trend where a significant effect of TrueSuccess enrollment begins at the fidelity measure 12 threshold.

⁵ (1| x) is notation for random effects

Table 8: LME Model of TrueSuccess participation against math and reading CGI scores, N = 1,048

	<u>β</u>	<u>Error</u>	<u>df</u>	<u>t</u>	<u>p</u>
Math					
TS	0.22	0.09	454	2.38	*
Grade					
K	0.13	0.10	2050	1.31	-
2nd	-0.06	0.10	2048	-0.63	-
3rd	-0.50	0.09	2076	-5.28	***
4th	-0.68	0.09	2079	-7.28	***
5th	-0.47	0.10	2079	-4.87	***
FRL	-0.12	0.06	2038	-1.97	*
Gender	0.10	0.05	1917	1.78	-
Attendance	3.34	0.68	2053	4.91	***
Reading					
TS	0.20	0.09	367	2.29	*
Grade					
K	0.14	0.10	2052	1.40	-
2nd	-0.31	0.10	2027	-3.12	**
3rd	-0.03	0.09	2063	-0.39	-
4th	-0.11	0.09	2073	-1.22	-
5th	-0.20	0.09	2080	-2.12	*
FRL	-0.04	0.06	2070	-0.71	-
Gender	-0.08	0.05	1992	-1.48	-
Attendance	1.89	0.65	2075	2.89	**

Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$

Error = standard error, TS = TrueSuccess

Reference group for grade is “1st”

Reference group for gender is “Female”

Table 9: Results indicate significantly improved math and reading CGI scores among TrueSuccess participants, N = 1,048

	<u>MSE</u>	<u>df₁</u>	<u>df₂</u>	<u>F</u>	<u>p</u>
Math					
TS	7.58	1	454	5.65	*
Grade	30.62	5	2033	22.82	***
FRL	5.18	1	2038	3.86	*
Gender	4.27	1	1917	3.18	-
Attendance	32.39	1	2053	24.13	***
Reading					
TS	5.09	1	367	5.27	*
Grade	4.79	5	2011	4.96	***
FRL	0.49	1	2070	0.50	-
Gender	2.10	1	1992	2.18	-
Attendance	8.05	1	2075	8.34	**

Note. - $p > .05$, $*p < .05$, $**p < .01$, $***p < .001$
 MSE = mean square error, df_1 = numerator degrees of freedom
 df_2 = denominator degrees of freedom, TS = TrueSuccess
 FRL = free/reduced lunch

Approach #3: Multiyear, Single Fidelity Threshold

Models for the multiyear analysis are the same as Model 1 and Model 2 with slight adjustments:

$$\text{Math/Read CGI} \sim \text{TSMulti} + \text{Grade} + \text{FRL} + \text{Gender} + \text{Attendance} + (1|\text{School}) \quad (3)$$

$$\text{Suspensions} \sim \text{TSMulti} + \text{Gender} + \text{Race/Ethnicity} + (1|\text{School}) \quad (4)$$

“TS” is replaced with “TSMulti”, a $k = 4$ categorical predictor with the following levels: control, 17-18 TS, 18-19 TS, and 2 year TS. These models were fit using only data from the 18-19 school year (i.e., the 2nd year of data). As such, (1|ID) becomes a singular measure and thus must be taken out. Finally, (1|School) is an added in random effect that significantly improves the model in the absence of the (1|ID) random effect. Model 3 is also used for reading.

Table 10 shows the results from both the LMERS for math and reading as well as the GLMM for suspensions. When the control group is set as the reference group, none of the individual group comparisons are significant for any measure. This is not due to lack of mean differences, but rather high variability of the measures and lower sample sizes. For example, math scores were equal to -0.168 ± 0.188 (5% margin of error) in the control group and -0.446 ± 0.240 in the 17-18 TS group, which when tested yields $p = .098$. This mean difference of 0.278 was non-significant, whereas the same comparison from the fidelity measure 12 assessment ($M_{CTRL} = -0.014 \pm 0.076$, $M_{TS} = 0.165 \pm 0.084$) was significant at $p = .022$ despite the lower mean difference of 0.179. Case sizes ($n_{1yr} = 1048$, $n_{2yr} = 256$) suggest insufficient power to make any proper conclusions from the multiyear data.

Table 10: Results indicate no significant differences at 2 years of participation, N = 266 per group

	<u>β</u>	<u>Error</u>	<u>df</u>	<u>t/z</u>	<u>p</u>
Math					
17-18	-0.21	0.22	113	-0.94	-
18-19	-0.33	0.20	127	-1.67	-
Both	0.08	0.22	104	0.37	-
Read					
17-18	0.07	0.24	108	0.31	-
18-19	-0.08	0.22	122	-0.36	-
Both	0.06	0.24	99	0.27	-
Susp					

	<u>β</u>	<u>Error</u>	<u>df</u>	<u>t/z</u>	<u>p</u>
17-18	0.12	0.49	-	0.24	-
18-19	0.67	0.44	-	1.52	-
Both	0.74	0.48	-	1.54	-

Note. - $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$

Error = standard error, df = degrees of freedom

Math/Reading LMER models use t

Suspension GLMM model uses z (no df)

Fidelity as a Factor

In order to test the effects of fidelity on the dependent measure, the unmatched data for all TrueSuccess students was modeled as follows:

$$\text{Math/Read CGI} \sim \text{Fidelity} + \text{Grade} + \text{FRL} + \text{Gender} + \text{Attendance} + (1|\text{School}) \quad (5)$$

$$\text{Suspensions} \sim \text{Fidelity} + \text{Gender} + \text{Race/Ethnicity} + (1|\text{School}) \quad (6)$$

Reading uses the same model as math (Model 5). “Fidelity” is the fidelity measure of the student’s teacher added in as a continuous predictor. Math significantly increases with fidelity, $\beta = 0.005$, $t(1,931) = 2.51$, $p < .05$, while suspensions significantly decrease with increasing fidelity, $\beta = -0.023$, $z = -5.82$, $p < .001$. Reading does not significantly change with fidelity, $t(1,941) = 0.04$, $p > .05$. The suspension coefficient when converted to a rate ratio is equal to 0.977. These results indicate that math scores increase by 0.005 per fidelity unit, while suspensions decrease by 2.3% per fidelity unit.

Appendix D: TrueSuccess School Participation

Table 11: TrueSuccess Participation by School

School	Ctrl, UM	TS, Fid 0	Ctrl, M	TS, Fid 12
Aberdeen Academy	266	0	25	0
Brookside Elementary	486	0	53	0
<i>Buchanan Elementary</i>	<i>111</i>	<i>498</i>	<i>13</i>	<i>205</i>
<i>Burton Elementary</i>	<i>602</i>	<i>234</i>	<i>100</i>	<i>22</i>
CA Frost Environmental Science Academy	867	0	33	0
Campus Elementary	579	0	52	0
<i>Cesar E. Chavez Elementary</i>	<i>249</i>	<i>424</i>	<i>44</i>	<i>329</i>
Coit Creative Arts Academy	430	0	38	0
Congress Elementary	428	0	30	0
Dickinson Academy	341	0	49	0
East Leonard Elementary	297	0	30	0
Gerald R. Ford Academic Center	413	0	38	0
Grand Rapids Montessori Academy	438	0	23	0
Harrison Park Academy	910	0	120	0
Ken-O-Sha Park Elementary	207	0	16	0
<i>Kent Hills Elementary</i>	<i>177</i>	<i>170</i>	<i>20</i>	<i>80</i>
Martin Luther King Jr. Leadership Academy	467	0	37	0
Mulick Park Elementary School	425	0	34	0
North Park Montessori Academy	606	0	30	0
Palmer Elementary	400	0	37	0
Ridgemoor Park Montessori	128	0	7	0
<i>Shawmut Hills Academy</i>	<i>158</i>	<i>157</i>	<i>17</i>	<i>66</i>
Sherwood Park Global Studies Academy	366	0	32	0
<i>Sibley Elementary</i>	<i>294</i>	<i>221</i>	<i>42</i>	<i>127</i>
Southwest Elementary - Academia Bilingue	643	0	122	0
<i>Stocking Elementary</i>	<i>113</i>	<i>264</i>	<i>6</i>	<i>219</i>

Note. UM = Unmatched, M = Matched.

Bold/Italic schools participated in TrueSuccess program.

Appendix E: TrueSuccess® Promotional Overview



MISSION

TrueSuccess® Inc. has a passion to help reverse the culture's character crisis by inspiring kids with character (social and emotional skills) to unleash their potential in school, work and life. TrueSuccess uniquely provides research-based (see summary page) educational tools that can simultaneously equip kids with social, emotional and literacy skills. (see Offering)

HISTORY OF SUCCESS

Since 2006, nonprofit TrueSuccess® Inc. has reached 200,000+ kids in grades K-12. The first seeds that grew to be TrueSuccess were sown in 1999. As his oldest of five children entered adolescent years, founder Bill Heneveld became acutely aware of the negative influence our media and culture can have on our kids' choices. Mr. Heneveld became passionate about equipping kids with positive behavior skills to make wise choices toward their best life. After seven years of groundwork TrueSuccess was formally founded in 2006 and is a 501(c)(3) nonprofit organization.

TrueSuccess has grown to provide grade-appropriate educational curriculum that can simultaneously equip kids, grades K-12, with positive behavior and literacy skills. The curriculum is flexible and uses have included in-classroom, after-school, mentoring, parenting, community centers and camps. TrueSuccess curriculum includes interactive tools that engage students with intriguing visuals and videos, inspiring real-life stories of positive role models and other fun, educational activities. As one student said, "TrueSuccess is educational and fun." The virtue themes of TrueSuccess include: Respect, Responsibility, Gratitude, Self-Control, Grit, Courage, Encouragement, Compassion, Integrity and Wisdom. With the continuing character-crisis in America, TrueSuccess continues to be passionate about significantly expanding the number of kids equipped with positive behavior skills to make a difference in the future of our country.

OFFERING

TrueSuccess achieves its mission by providing tools, training and support to improve kids' behavior and literacy skills through its full cycle SEL (Social and Emotional Learning) program (see visual). This full SEL cycle includes: 1) Curriculum/Tools (online and printed for grades K-12), 2) Educator Training/Professional Development, 3) Educator Coaching (year-long support), 4) Administrator Support (including fidelity of implementation Teacher-Use Analytics), 5) Assessments (for students and teachers), and 6) Adaptation Interventions (identifying TrueSuccess curriculum to address assessed deficiencies).

Curriculum is organized around monthly virtues/positive behavior skills; Respect, Responsibility, Gratitude, Self-Control, Grit, Courage, Encouragement, Compassion, Integrity and Wisdom. Curriculum is developed fresh every school year with new content and lessons. Website SmartCharts provide easily accessible lessons and tools that include well over 1,000 activities. At the core of TrueSuccess are educator-friendly No-Prep Lessons which include virtue videos, lifestories, virtue activity worksheets, literacy exercises and many other activities. Lifestories are a very effective curriculum component that

consist of engaging biographies of positive role models whose lifestories of overcoming adversity inspire kids to become superheroes in their own lifestories. Teachers find TrueSuccess very easy to implement (96%) and see high value to continue to do TrueSuccess (85%).

© 2021 TrueSuccess Inc. ♦ info@TrueSuccessTools.org ♦ 616.977.5900 ♦ Grand Rapids, MI 49546 ♦ www.TrueSuccessTools.org

Appendix F: Bibliography

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

<https://doi.org/10.18637/jss.v067.i01>

Ho, D. E, Imai K., King, G., Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1–28.

<https://doi.org/10.18637/jss.v042.i08>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.

<https://doi.org/10.18637/jss.v082.i13>

NWEA 2020 MAP Growth: Achievement Status and Growth Norms for Students and Schools. Meng Thumb, Yeow; Kuhfeld, Megan. July, 2020.

<https://teach.mapnwea.org/impl/normsResearchStudy.pdf>

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>