The Impact of Eight:
How Eight or More Skills in Exact Path Led to Gains on State Achievement Tests

Edmentum Efficacy Research & Learning Engineering

Aaron Butler, PhD
Weiling Li, PhD
Amy J. Dray, EdD

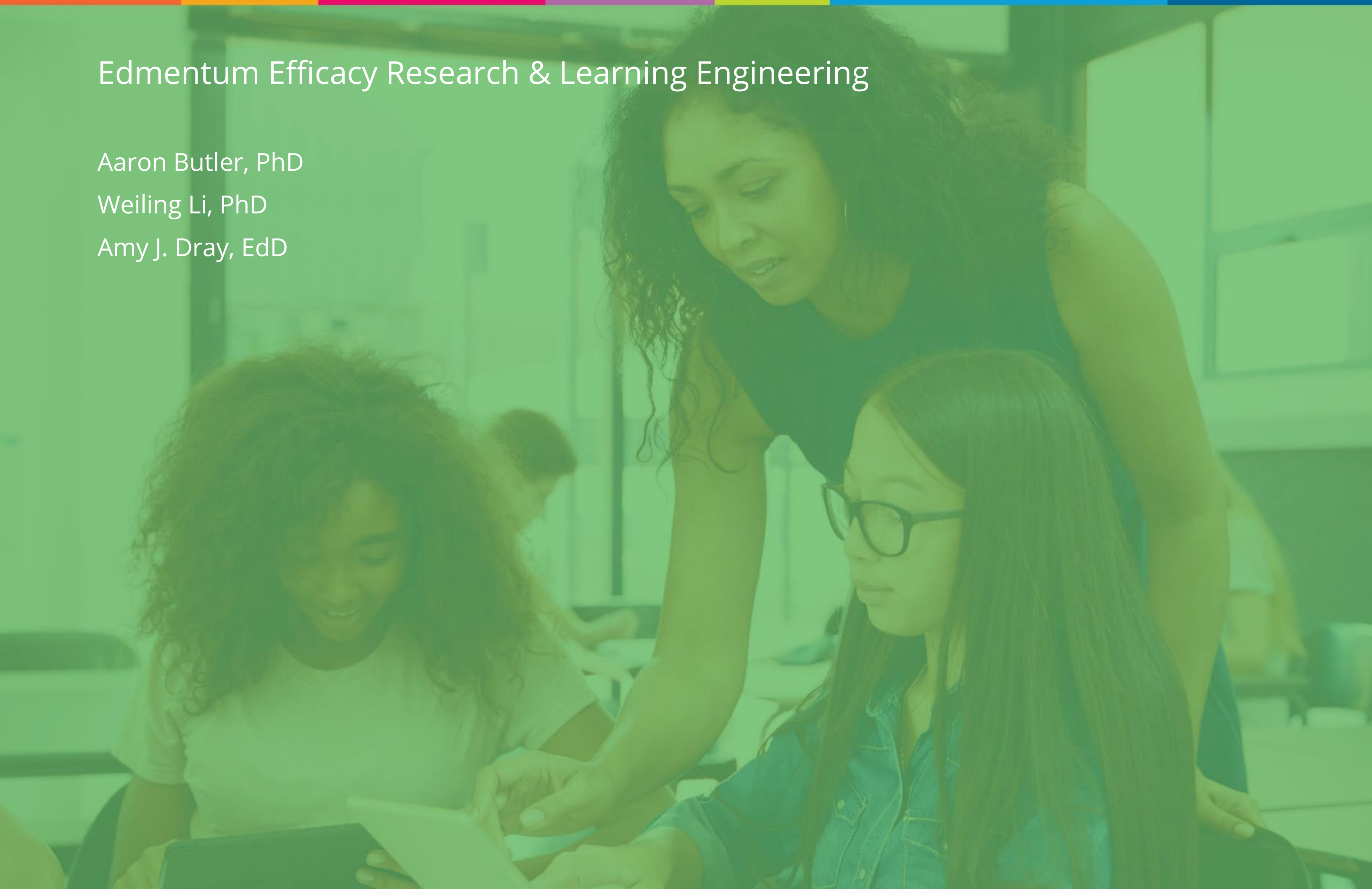

Table of Contents

Customer Support
800.447.5286
support@edmentum.com
05.31.23

Abstract

Edmentum offers a diagnostic-driven, individualized instruction program called Exact Path, designed to propel academic growth and achievement for K–12 students in math, reading, and language arts. In this 2023 study, we evaluated the efficacy of Exact Path on student achievement in mathematics.

Results of this study confirmed that the completion of 8+ skills in Exact Path corresponded to a significant and sizeable improved achievement on state assessment results in math with improvement indexes as high as 19 percentile points. When Exact Path is implemented with fidelity, an average student who used the program could accelerate learning for a gain of up to 19 percentile points as compared to a student who did not participate in the program.

The improvement index is a reframing of effect size, a statistical indication of the magnitude of impact on student achievement and a critical measure for understanding how well educational interventions support student success. The effects of using Exact Path found in this rigorous study were of practical as well as statistical importance. The effects exceeded or matched the average effects (e.g., 0.16) of nearly 750 rigorous studies of interventions intended to improve mathematics and reading achievement[1]. Thus, teachers should feel confident using the personalized learning of Exact Path alongside core curricula.

Since 2017, almost 1,300 schools and districts in Arizona have used Exact Path. These locations include districts as diverse as Phoenix, where multiple schools participated, as well as smaller locations such as Pinon. This study evaluated the efficacy of Exact Path in a large urban district so that we can provide specific recommendations to state-specific educators and inform the broader community of policymakers and practitioners about personalized learning usage that may support increased student academic achievement.

The rigorous quasi-experimental design of the efficacy study meets the requirements for **ESSA Tier 2 evidence** and the **What Works Clearinghouse 5.0 Group Design standards with reservations**.

[1] Kraft, M. (2020). Interpreting effect sizes of educational interventions. *Educational Researcher, 49(4)*, 241-253.

Customer Support
800.447.5286
support@edmentum.com

Rationale

Historically, underserved students have been less likely to be assigned to qualified educators and have lacked access to high-quality courses. Low-income students, students in rural communities, and students of color are more likely to attend schools that receive lower funding levels and offer less educational opportunity than more affluent students (Berliner, 2008; Goldhaber, Quince, & Theobald, 2018). In Arizona, where 45 percent of K–12 students identify as Latino/Hispanic, educational disparities are notable (Arizona Department of Education, 2021). Yet, resource allocation is key to understanding academic outcomes; the mechanisms underlying successful use of school investments are not well understood (Handel & Hanushek, 2022).

Personalized learning has been proposed as a solution in states such as Arizona where access to quality education is uneven (Pane, Steiner, Baird, & Hamilton, 2015). But there is much to learn following the recent pandemic about the efficacy of technology education (Goldhaber et al., 2022). The concept of personalized learning has existed for some time, but the adoption of personalized learning in schools increased significantly during COVID-19 when many school districts had to shift to remote teaching. The sophistication of current technology platforms, digital content, and computer-adaptive testing approaches has also increased over time, providing learners with educational opportunities that were not previously possible.

While personalized learning is a general term, for this project we suggest it is composed of three interrelated concepts. First, a personalized trajectory of learning in a virtual school setting should be grounded in the learning progression of specific disciplinary knowledge, such as math or reading (National Research Council, 2001). The underlying content of what a student should learn, and how that content advances over time, should be the same online as in a traditional curriculum, because the learning progression provides a roadmap for instruction and must be aligned with state standards (Valencia, Pearson, & Wixon, 2014). Second, personalized learning accommodates, and provides access to, individual learning paths where students progress through a program of instruction that meets their needs, whether these needs are remedial, grade-level instruction, or enrichment (Means, Bakia, & Murphy, 2014; Pane et al., 2015). An online program may provide instructional flexibility. Third, for a learning path to be truly individualized, the person needs to be fairly, accurately assessed at the onset of their learning so that their location on the underlying learning progression is captured accurately. Assessment provides guidance for instruction; personalized learning platforms typically include an algorithm to recommend where each student should begin their journey in learning progression, such that the instruction students receive is optimally suited to their current achievement level.

Edmentum has provided technology-based learning solutions for over 60 years. Combining advances in education with learning engineering and psychometrics, the Exact Path curriculum offers instruction in math, reading, and language arts. It is grade agnostic, meaning that the learning path offered to students depends on their performance on an initial assessment. Learning paths accommodate students who are still struggling with grade-level precursor skills as well as students who would be best served by above-grade-level enrichment opportunities. The initial assessment is typically the Exact Path diagnostic assessment, which was developed by Edmentum. Following initial assessment and placement into a learning path, a student moves through their learning trajectory and is further assessed at key touchpoints via computer-adaptive tests. After each assessment, the learning path is further refined based on the student's current level of content knowledge.

As researchers embedded in the organization, we wished to examine the quality of initial assessment at designing a learning path and to better understand how Exact Path usage promotes positive student outcomes. The following research question guided the design and analyses used in this study.

- *To what extent, if any, does Exact Path usage impact student achievement outcomes in math as measured by Arizona state test scores?*

Methods

Data and Sample

We used student and course-level data from a suburban public school district in Arizona in our study. The district provided data through a data-sharing agreement and agreed to participate anonymously in this study. The data contained demographic and Arizona state test information on 2,435 students in grades 3–8 for school years 2020–2021 and 2021–2022.

This study of Exact Path began September 26, 2022, and is scheduled to be completed June 30, 2023. We observed students between the state testing windows. For 2020–2021, those windows for grades 3–8 were as follows: for the computer test, April 5–30 with writing completed by April 16; for paper-based tests, April 5–14. In 2021–2022, the windows for grades 3–8 were as follows: for computer-based tests, April 4–29 with all writing completed by April 15, 2022; for paper-based tests April 4–13.

All public school students in the state were required to take the math state test unless their caregivers opted out (Cambium Assessment, 2021), and this was our population of interest. Our analytic sample consisted of all the students in the sampling frame for whom the Exact Path curriculum was made available by instructor choice, and students who have state test scores in both the 2020–2021 and 2021–2022 school years. All the students in the sample took the Exact Path diagnostic assessment and were assigned a learning path. Although the district made the curriculum available, classroom implementation is up to the school and teacher. Student information was provided by the district; school and class-level information were not available.

Table 1. Demographic Characteristics of the Sample, Number, and Proportion per Student (n=4,806)

Demographic Characteristics	Math (n=2,420)
Free/reduced lunch	2,112 (87.2%)
Female	1,158 (47.9%)
Male	1,262 (52.1%)
American Indian - Alaska Native	35 (1.4%)
Asian	12 (0.5%)
Black - African American	98 (4.0%)
Hispanic or Latino	2,122 (87.7%)
Pacific Islander	11 (0.5%)
White	142 (5.9%)
Special education	247 (10.2%)
English language learner	506 (20.9%)

Customer Support
800.447.5286
support@edmentum.com
05.31.23

Research Design

The study used a nonrandomized control group, pretest-posttest quasi-experimental design. The design meets What Works Clearinghouse (WWC) 5.0 standards with reservations (WWC, 2022). According to the WWC, a quasi-experimental design (QED) uses a non-random process to form the intervention and comparison conditions. The WWC allows groups to be formed using a variety of methods as long as the groups are mutually exclusive. That is, units (e.g., students or schools) can only be analyzed as a member of a single group. Further, in a QED, the WWC accepts assignment to the intervention based on observed characteristics. Assignment to study conditions for this study was conducted at the student level.

Propensity Score Weighting

In this study of Exact Path, we used student-level data from the school district to assess the pre-to-post Arizona state assessment achievement outcomes for Exact Path students relative to a control group of students. We used Inverse Probability-of-Treatment Weighting (IPTW) Propensity Score methods (Austin & Stuart, 2015) to estimate causal effects. Propensity score weighting is used to isolate the effect of an intervention from other differences that may exist between the intervention and control groups. Propensity score weighting assigns students different weights—weighting them up or down to make the students in the intervention group and the control group more similar to each other. Weighting subjects by the inverse probability of intervention received creates a synthetic sample in which assignment to an intervention is independent of measured baseline covariates. IPTW using the propensity score allows us to obtain unbiased estimates of average intervention effects (Austin & Stuart, 2015). Additionally, an advantage of using propensity score weighting, as opposed to matching, is that it allowed us to leverage information from all students in the analytic sample. The process by which we used IPTW is presented in Appendix B. IPTW was conducted within grades and subjects. The WWC considers IPTW an appropriate method to achieve baseline equivalence between intervention and control groups (WWC, 2022).

Intervention/Control Groups

We defined the intervention group as students who had both 2020–2021 and 2021–2022 Arizona state assessment scores and completed at least eight Exact Path lessons in math. That is, students needed to complete at least eight lessons within the seven sub-domains of math (Algebra & Expressions; Counting & Cardinality; Fractions & Ratios; Functions; Geometry; Measurement, Data, & Statistics; and Numbers & Operations). At least eight lessons were chosen as the definition for Exact Path based on prior research (Randel, 2018a; 2018b) and substantive understanding of the Exact Path curriculum. For example, Exact Path assigns lessons in groups of three to four. Using eight lessons helps ensure that students are working their way through the learning progression and are using Exact Path as intended. Students are expected to complete a set of lessons, take a progress check, and move further along the learning progression. Also, up to 31 skills are available for math. This means 12 lessons represent approximately one semester's worth of learning on the learning progression. Since the study examined student achievement from spring 2021 to spring 2022, a minimum of eight lessons was deemed to be a reasonable definition of Exact Path use. The number of lessons completed by students in the Exact Path intervention group varied (see Appendix A for details).

We defined the control group as students who had both 2020–2021 and 2021–2022 Arizona state test scores and completed fewer than eight lessons in math. This definition helped ensure that students in the control group were not using Exact Path as intended: to address weaknesses in their math achievement as identified by the diagnostic assessment. This definition of the control group also ensured that no students were included in both groups. In other words, the study groups were mutually exclusive.

Measures

The outcome measures for the study are the Spring 2022 Arizona state tests in math (Cambium Assessment, 2021). The reliability coefficients for all subjects and grades for the test ranged from 0.90 to 0.94. The tests were designed to measure student progress toward achievement of the Arizona State Standards. As a criterion-referenced system of tests, the meaning of test scores was critically evaluated by the degree to which test content was aligned with the Arizona State Standards.

The interim assessment, the Exact Path diagnostic assessment, was used in the study as the mechanism by which students are assigned to a personalized learning path in the Exact Path curriculum. It is not used as a variable in this study because it was designed by the curriculum developer. The technical specifications of the assessment including the field test are reported elsewhere in the Adoption list submission.

The Exact Path diagnostic assessment provides multiple types of feedback to help students and teachers understand progress throughout the year. Specifically, Exact Path delivers scale scores that can be used to see where students start (the scale score on their first test) and growth measures to quantify how much student scores change. Growth can be calculated between two tests, fall to winter, winter to spring, or fall to spring. Additionally, Exact Path provides Norm Referenced Scores/Percentiles relative to the end of the school year. Students, schools, and teachers can consult the assessment dashboard as the year progresses to see performance over the year. Moreover, Exact Path offers performance levels called Grade Level Proficiency classifications in score reports. These classifications can also be used to understand progress. Finally, Exact Path provides both Lexile and Quantile measures upon completion of the diagnostic assessment.

For teachers and administrators, interpretation guides and workbooks for data-driven instruction are available as well as digital resources and videos on important assessment literacy topics such as “Unpacking National Percentile Rank” and “Making Sense of Scale Scores.” Videos ensure that teachers with differing literacy skills can take appropriate action upon receiving information from the interim assessment. Furthermore, Edmentum provides resources developed specifically for educators and families. These resources span the Exact Path experience—from introducing the purpose of the program to interpreting assessment results, understanding learning best practices, and accessing instructional examples. Additional examples of the Exact Path diagnostic assessment interpretive resources described here can be found elsewhere in the Adoption list submission.

Independent Variables

We used the students’ Arizona state test scores from spring 2021 as a pretest measure of students’ academic achievement before they enrolled in Exact Path in fall 2022. Additionally, demographic information such as gender, ethnicity, free/reduced lunch status (SES), English language learner (ELL) status, and grade-level information were captured. Gender was dichotomously coded (1=Female, 0=Male). Race/ethnicity indicators were coded as follows: Asian (1=Yes, 0=No), Hispanic (1=Yes, 0=No), Black (1=Yes, 0=No), Pacific Islander (1=Yes, 0=No), and White (1=Yes, 0=No). Socioeconomic status (SES) was determined by proxy of free or reduced-price lunch (1=Yes, 0=No). English language learner (ELL) status was dichotomously coded (1=Yes, 0=No).

Analytic Procedures

First, descriptive statistics and correlations were examined along with linearity. Subsequently, a series of weighted regression models were fit to the data. Residuals were examined to investigate our assumptions. Weighted least squares estimation is equivalent to maximum likelihood estimation of β with independent errors with variance inversely proportional to the weights. Points with high weights have low error variances and are thus expected to lie closer to the fitted regression line (Gelman & Hill, 2007).

The average intervention effect of Exact Path instructional usage on student achievement was estimated by calculating the differences between intervention and control groups on the 2021 (pretest) and 2022 (posttest) Arizona state scores using regression analysis. We used inverse probability of treatment weights and included the covariates as additional controls in our models. This approach is considered "doubly robust," because they include two mechanisms for removing selection bias (Kang & Schafer, 2007).

We conducted our impact analyses using the following linear regression model fit to the data separately for each grade level and subject:

$$Y_i = \beta_0 + \beta_1 Intervention_i + \beta_3 \boldsymbol{X} + \varepsilon_i,$$

where Y_i were the standardized scaled scores on Arizona state assessment in math for student i in 2021–2022. $Intervention_i$ was a binary variable indicating whether student i was an Exact Path user of 8 or more skills (1=Yes, 0=No). $\boldsymbol{X}$ was a vector of student characteristics, including gender, race/ethnicity, eligibility for the school lunch program, special education status, English language learner status, and 2021 Arizona state test scores (standardized); ε_i represented the model's random error.

Weight estimates, using propensity scores ($\hat{e}(x)$), for students in the intervention group were calculated using the following equation:

$$\omega = \frac{1}{\hat{e}(x)}.$$

And for students in the control group, the following equation was used:

$$\omega = \frac{1}{1 - \hat{e}(x)}.$$

Results

Use of Exact Path Positively Impacted Math Achievement

As reported in Table 2, we found statistically significant positive impacts of Exact Path usage (intervention) on math achievement for all grades. For example, controlling for other variables in the model, the effect size of Exact Path for grade 4 was 0.45 ($p < 0.01$). In terms of practical significance, that translates into an improvement index of +18, showing the expected change in percentile rank if a comparison student had received the intervention. A student at the 50th percentile at pretest, for example, could be expected to shift into the 68th percentile. In grade 8, the effect of Exact Path is lessened (0.145), so it could be expected that a comparison student would shift 6 percentile points—from the 50th to the 56th percentile—had they received the intervention.

In Table 2, we can see that while the effect of Exact Path is positive across all the grades, the magnitude of the effect decreases from grade 4 (0.45) to grade 8 (0.15). According to Bloom et al. (2007), annual

edmentum
Customer Support
800.447.5286
support@edmentum.com
05.31.23

gains in the early elementary grades are usually followed by a gradual decline in the later grades. A given intervention, therefore, will vary across grades, and this study supports Bloom et al.'s (2007) findings.

Table 2. Effects of Exact Path Usage on Math Achievement, Grades 4–8 (n=2,420)

Grade	**Grade 4 (*n*=484)**	**Grade 5 (*n*=463)**	**Grade 6 (*n*=454)**	**Grade 7 (*n*=479)**	**Grade 8 (*n*=540)**
Exact Path (>=8 skills)	0.452***	0.483***	0.315***	0.189***	0.145**
	(0.051)	(0.055)	(0.058)	(0.055)	(0.056)
Prior year test score	0.646***	0.658***	0.641***	0.676***	0.770***
	(0.029)	(0.031)	(0.039)	(0.032)	(0.032)
Male	-0.161**	0.093	0.022	0.127*	-0.064
	(0.052)	(0.057)	(0.059)	(0.057)	(0.057)
Asian	-0.241	-0.241		0.170	1.149
	(0.459)	(0.600)		(0.497)	(0.953)
Black	-0.141	-0.213	-0.236	0.226	0.252
	(0.291)	(0.492)	(0.342)	(0.217)	(0.291)
Hispanic	-0.242	-0.484	0.041	0.119	0.401
	(0.234)	(0.475)	(0.318)	(0.152)	(0.265)
Pacific Islander	-0.195	-0.797	0.096	-0.413	
	(0.603)	(0.670)	(0.411)	(0.866)	
White	0.016	-0.292	0.403	0.499*	0.402
	(0.244)	(0.487)	(0.360)	(0.198)	(0.288)
Socioeconomic status	-0.078	-0.122	-0.283**	-0.244**	-0.196*
	(0.087)	(0.080)	(0.094)	(0.083)	(0.095)
Special education	-0.389***	-0.033	0.002	-0.301**	0.278**
	(0.096)	(0.092)	(0.108)	(0.094)	(0.103)
English learner	-0.086	-0.143	-0.268***	-0.078	-0.088
	(0.070)	(0.076)	(0.075)	(0.078)	(0.077)
Intercept	-0.014	0.080	-0.016	-0.114	-0.305
	(0.240)	(0.483)	(0.331)	(0.177)	(0.281)
R2	0.667	0.634	0.562	0.619	0.608
R2 Adj.	0.660	0.625	0.552	0.610	0.600

Notes: Standard errors in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Customer Support
800.447.5286
support@edmentum.com

Limitations

There are a few limitations to this study. First, this study was not an experimental research study with Exact Path assigned randomly to students, thus causality cannot be fully inferred from the study's results. As a result, we chose to examine the study's research questions using a quasi-experimental design that meets the What Works Clearinghouse 5.0 standards with reservations (WWC, 2022). A key limitation of propensity score matching techniques such as IPTW is their inability to account for potential bias stemming from unobservable covariates (Austin & Stuart, 2015). Therefore, it was difficult to know whether the assumption of ignorable treatment assignment had been met (Rosenbaum & Rubin, 1983). Austin and Stuart (2015) recommended that studies employing propensity score matching methods include all available (observable) covariates in the model used to estimate propensity scores, as we did in our models used to estimate propensity scores.

Additionally, the district in which this study was conducted uses Exact Path extensively. While this makes the district a good site for studying the curriculum, future research should incorporate districts with zero usage and/or classrooms for which Exact Path was not an option. It would be worthwhile to compare results between curriculum users, non-compliers, and "never takers." Additionally, although we know the district uses the curriculum, future research should examine classroom variation in implementation and fidelity. Also, the study does not account for the multi-level structure of the data (the clustering of students within schools). Further research may address this limitation by examining how the program's impact varies across schools and by combining districts with high and low usage to examine variation between them. Future research also should re-examine the methodology for the subgroup analyses because, while it is informative to examine heterogeneity, other analytical techniques may be preferable.

Conclusions

We found Exact Path usage to have a positive impact on math achievement as measured by the Arizona state test, and the effects were statistically significant for grades 4–8. The effect was captured between the spring 2021 and spring 2022 Arizona state test windows, during which the students in the sample took the Exact Path diagnostic assessment, which was used to generate individual learning path trajectories in the Exact Path program, and which also provided targeted information to guide instruction. Students participating in the Exact Path intervention showed greater gains in math achievement than students in the control group.

The statistically significant gains made by students in the Exact Path math intervention group suggest that Exact Path lessons are supporting learning for the skills students need to develop in order to improve their achievement. Had Exact Path targeted skills students had already mastered, it is likely students would not have seen the same gains in achievement between administrations of the state test. These results suggest a practical importance of Exact Path usage and completion of at least eight lessons as being an optimal target of instruction.

Our results are generalizable to the states of Arizona and South Carolina, because the intervention is likely to produce positive results in different circumstances. For example, for all ELL students in the sample, the treatment versus comparison effect size on math achievement is 0.17; for all female students, the effect size is 0.27; for all free/reduced price lunch students, the effect size is 0.33, and so forth. The positive effects suggest the intervention flexibly adapts to the characteristics of students who

edmentum

Customer Support
800.447.5286
support@edmentum.com
05.31.23

were assigned to learning paths by the Exact Path diagnostic assessment. High adaptability suggests that Exact Path is generalizable to a variety of educational contexts.

The data on lesson completion in Appendix B show that many students in the Exact Path intervention group completed more than eight lessons. In every grade except grade 4, the majority of students completed between 8 and 12 lessons. As prior research has shown, Exact Path usage is positively correlated with achievement as measured by the diagnostic assessment (Edmentum, 2017; 2018). Although not addressed in this study, an increase in the number of Exact Path lessons completed likely results in increases in scores on the diagnostic assessments.

This study was conducted at the level of rigor needed to meet WWC 5.0 standards with reservations (WWC, 2022). Baseline equivalence was established. The measure used to establish baseline equivalence and as the math achievement outcome meets WWC standards for validity and reliability. The baseline and outcome measures are aligned to national and state academic content standards and so are not over-aligned to the Exact Path intervention. The study had no observable confounds.

The study also meets criteria set forth by the Every Student Succeeds Act (U.S. Department of Education, 2016). The Department of Education considers a quasi-experimental study to be "well-designed and well-implemented" if it receives a Meets WWC Design Standards with Reservations rating or is of equal quality (U.S. Department of Education, 2016). The study also meets the ESSA criteria for statistically significant positive effects. These two aspects of the study mean it qualifies as providing Moderate Evidence (Level 2) of Exact Path's effectiveness.

References

Arizona Department of Education. (2021). *Four year graduation rate 2015-2021.* https://www.azed.gov/accountability-research/data/

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using Inverse Probability of Treatment Weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Berliner, D. C., (2006). Our impoverished view of educational research. *Teachers College Record, 108,* 949-995. https://doi.org/10.1111/j.1467-9620.2006.00682.x

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328. https://doi.org/10.1080/19345740802400072

Cambium Assessment. (2021). *Annual technical report: Arizona State Assessments for English Language Arts and Mathematics.* Arizona Department of Education. https://www.azed.gov/sites/default/files/2022/02/SY2020-21_AzMERIT_Annual%20Technical%20Report.pdf

Edmentum. (2017). *Edmentum Exact Path – Research base for math: Growth through adaptive assessments and learning paths.* Edmentum. https://cdn.edmentum.com/edmentumcity/edmentum2/uploads/2018/08/Edmentum-Exact-Path-Research-Base-for-Math-Research-Paper.pdf

Handel, D., & Hanushek, E. A. (2022). *U.S. School and Finance: Resource and outcomes. Working Paper 30769.* National Bureau of Economic Research. https://www.nber.org/papers/w30769

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological methods*, *15*(3), 234–249. https://doi.org/10.1037/a0019623

Hirano, K., Imbens, G., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica, 71*(4), 1161–1189. https://doi.org/10.1111/1468-0262.00442

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2022). The consequences of remote and hybrid instruction during the pandemic. Working Paper No. 267-0522. CALDER. https://caldercenter.org/sites/default/files/CALDER%20Working%20Paper%20267-0522_0.pdf

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in US public schools. *American Educational Research Journal, 55*(1), 171–20. https://doi.org/10.3102/0002831217733445

Kang, J. D. Y. & Shafer J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population pean from incomplete data. *Statistical Science 22*(4), 523–539. 10.1214/07-STS227

Markoulidakis, A., Taiyari, K., Holmans, P. et al. A tutorial comparing different covariate balancing methods with an application evaluating the causal effects of substance use treatment programs for adolescents. *Health Serv Outcomes Res Method (2022)*. https://doi.org/10.1007/s10742-022-00280-0

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when, and how.* Routledge Taylor & Frances.

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031–1039. https://doi.org/10.1111/j.1469-7610.2010.02254.x

National Research Council. 2001. Knowing what students know: *The science and design of educational assessment.* Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. National Academy Press.

Pane, J. F., Steiner, E. D., Baird, M. D., & Hamilton, L. S. (2015). Continued Progress Executive Summary. In *Continued Progress: Promising Evidence on Personalized Learning: Executive Summary* (pp. 1–8). RAND Corporation. http://www.jstor.org/stable/10.7249/j.ctt19w724m.1

Randel, B. (2018a). *Impacts of Edmentum's Exact Path on student mathematics achievement.* Century Analytics. https://files.eric.ed.gov/fulltext/ED605132.pdf

Randel, B. (2018b). *Impacts of Edmentum's Exact Path on student reading achievement.* Century Analytics. https://files.eric.ed.gov/fulltext/ED605131.pdf

Rosenbaum, P. R., & Rubin, D. R. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. https://doi.org/10.1093/biomet/70.1.41

U.S. Department of Education, Office of Elementary and Secondary Education. (2016, September). Non-regulatory guidance: Using evidence to strengthen education investments. https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf

Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting Text Complexity in Context: Refocusing on

Comprehension of Complex Text. *The Elementary School Journal*, *115*(2), 270–289. https://doi.org/10.1086/678296

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2022). What Works Clearinghouse: Standards handbook (Version 5.0). http://whatworks.ed.gov

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4–12. https://doi.org/10.1111/j.1745-3992.2005.00014.x

Customer Support
800.447.5286
support@edmentum.com

Appendix A

Explanation of Procedures for Inverse Probability of Treatment Weighting Methods

We used Inverse Probability-of-Treatment Weighting Propensity Score to estimate the effects of Exact Path usage on student achievement. This semi-parametric approach gives us a better chance to isolate causal factors by removing selection biases associated with differences between students who use Exact Path instructional materials (treatment) and those who do not (comparison). While there are numerous approaches to propensity scores, here we use IPTW (as opposed to propensity score matching, stratification, and adjustment) given the increased practical applications that have emerged due to its benefits (Markoulidakis et al., 2022). It weights the data such that the treatment and control groups have nearly identical univariate distributions on all observed covariates (Austin & Stuart, 2015). IPTW has been shown to remove selection biases that work through factors associated with the covariates, yielding efficient estimates of the average treatment effects (Hirano, Imbens, & Ridder, 2003). Moreover, unlike propensity score matching, which defines particular cases from both groups to compare, IPTW preserves external validity and ensures that all respondents in the sample can be used.

We implemented IPTW in two steps. First, we determined the probability for each student to complete at least eight Exact Path lessons within math. Second, in comparing the achievement outcomes for Exact Path users with those who did not use the curriculum, we weighted each control case i such that the $weight_i = p_i/(1 - p_i)$, where p is the estimated probability of completing eight or more Exact Path lessons. This approach gives more weight to students who resemble Exact Path users and less weight to students who do not. As a result, we effectively compare two (weighted) groups that are more similar in their pre-treatment characteristics.

In estimating the propensity scores to be used in weighting, we selected only covariates that described pre-treatment conditions. Covariates included several demographic and socioeconomic variables known to relate to student achievement. Specifically, we used student gender (“female,” “male”), race/ethnicity (“Asian,” “Black/African American,” “Hispanic/Latino,” “Native Hawaiian/Other Pacific Islander,” “White”), special education status (“Yes,” “No”), English language learner status (“Yes,” “No”), and free or reduced price lunch status (“Yes,” “No”). We examined all covariates for missing data and out-of-range values prior to modeling propensity scores. No students were excluded from the analysis due to having covariates with missing or out-of-range values.

While there are several ways to balance covariates, here we calculated a standardized difference in means to use as a numerical balancing diagnostic to test whether covariate balance is achieved (Harder, Stuart, & Anthony, 2010). The smallest standardized difference indicates balance.

The results of the table show that sufficient balance was achieved for the weighted sample per the recommendations by Harder et al. (2010). For instance, in the table below, we are examining the divergence between the two groups for State Test Score 2021 (0.26) and how it adjusts to -0.033.

edmentum
Customer Support
800.447.5286
support@edmentum.com
05.31.23

Appendix B

Covariate Balance Tables

Table B1. Standardized Mean Difference of Unadjusted and Adjusted Covariates, Grade 4

Covariates	Math	
	SMD Unadjusted	SMD Adjusted
State Test Score 2021	0.266	-0.033
American Indian or Alaska Native	-0.045	0.012
Asian	0.009	0.008
Black or African American	-0.039	0.007
Hispanic or Latino	0.059	0.024
Native Hawaiian or Other Pacific Islander	0.004	0.004
White	0.012	-0.055
Special education	-0.091	0.032
English learner	0.063	0.079
Free/reduced price lunch	0.197	-0.036

Table B2. Standardized Mean Difference of Unadjusted and Adjusted Covariates, Grade 5

Covariates	Math	
	SMD Unadjusted	SMD Adjusted
State Test Score 2021	0.007	0.006
American Indian or Alaska Native	0.011	0.011
Asian	-0.085	-0.007
Black or African American	0.189	-0.019
Hispanic or Latino	0.007	0.006
Native Hawaiian or Other Pacific Islander	-0.129	0.002
White	-0.087	-0.010
Special education	0.023	-0.012
English learner	0.051	0.020

edmentum
Customer Support
800.447.5286
support@edmentum.com
05.31.23

Free/reduced price lunch	0.007	0.006

Table B3. Standardized Mean Difference of Unadjusted and Adjusted Covariates, Grade 6

Covariates	Math	
	SMD Unadjusted	SMD Adjusted
State Test Score 2021	0.381	0.145
American Indian or Alaska Native	-0.014	0.000
Asian	-0.043	0.002
Black or African American	0.105	0.002
Hispanic or Latino	-0.036	-0.001
Native Hawaiian or Other Pacific Islander	-0.012	-0.003
White	-0.061	0.005
Special education	0.001	-0.007
English learner	-0.002	0.000
Free/reduced price lunch	0.381	0.145

Table B4. Standardized Mean Difference of Unadjusted and Adjusted Covariates, Grade 7

Covariates	Math	
	SMD Unadjusted	SMD Adjusted
State Test Score 2021	0.339	-0.029
American Indian or Alaska Native	0.004	-0.019
Asian	-0.010	0.000
Black or African American	0.002	0.009
Hispanic or Latino	0.026	0.001
Native Hawaiian or Other Pacific Islander	0.003	0.002
White	-0.024	0.007
Special education	-0.044	-0.004
English learner	-0.094	-0.004
Free/reduced price lunch	0.047	-0.002

Table B5. Standardized Mean Difference of Unadjusted and Adjusted Covariates, Grade 8

Covariates	Math	
	SMD Unadjusted	SMD Adjusted
State Test Score 2021	0.117	-0.008
American Indian or Alaska Native	0.005	0.000
Asian	0.002	0.002
Black or African American	-0.005	-0.003
Hispanic or Latino	0.056	0.006
Native Hawaiian or Other Pacific Islander	-0.058	-0.006
White	0.006	0.013
Special education	-0.032	0.001
English learner	-0.046	-0.015
Free/reduced price lunch	0.117	-0.008

Customer Support
800.447.5286
support@edmentum.com

Appendix C

Exact Path Lessons Completed

Table C1. Number of Students in the Intervention Group Completing Math Lessons by Grade

Lessons completed	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
8 lessons	5	3	8	12	33	32
9 lessons	0	0	0	0	0	0
10 lessons	2	9	8	14	22	30
11 lessons	1	1	1	0	0	0
12 lessons	5	9	7	26	24	33
13 lessons	0	0	1	0	0	0
14 lessons	9	11	5	10	17	19
15 lessons	1	0	0	0	0	0
16 lessons	3	11	10	14	25	38
17 lessons	0	0	0	0	1	0
18 lessons	5	12	14	30	27	35
19 lessons	0	0	0	0	0	0
20 lessons	7	16	18	24	27	31
21 or more lessons	517	493	476	352	298	250
Total	555	565	548	482	474	468

Customer Support
800.447.5286
support@edmentum.com
05.31.23

Appendix D

Correlation Table

Table D1. Correlations Between State Test Scores, Treatment Status, and Covariates, Grades 4–8, Math (n=2,420)

Variable	1	2	3	4	5	6	7	8
1. Treatment	1							
2. Gender	-.02	1						
3. Ethnicity	-.03	.01	1					
4. Special education	-.03	.10***	-.02	1				
5. English learner	-.01	.03	-.05*	.09***	1			
6. Free/reduced lunch	.01	.01	-.16***	-.06**	.04*	1		
7. State test score, 2021	.07**	.03	.15***	-.26***	-.36***	-.11***	1	
8. State test score, 2022	.11***	.04	.15***	-.24***	-.33***	-.12***	.76***	1

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Customer Support
800.447.5286
support@edmentum.com
05.31.23