Research Report

How a Great City School District Is Improving Performance and Closing Achievement Gaps for All Students: A 10,000-Student Study of Des Moines Public Schools

Author
Lindsey Devers Basileo, PhD
Principal Investigator

For more of the consulting team members'* historical work visit InstructionalEmpowerment.com

*See Appendix for list of participating team members.

SUMMARY: Improving performance and closing achievement gaps

- Des Moines Public Schools (DMPS) and a team of external consultants partnered to establish demonstration schools that focused on building capacity of leadership, rigorous instruction and student academic teams at 22 schools.
- All demonstration school students gained the equivalent of an additional 11 days of learning in reading and six days in math.
- Year-two demonstration school students gained the equivalent of an additional 21 days of learning in reading and 10 days in math.
- All high school demonstration school students gained the equivalent of an additional 61 days of learning in reading and 42 days in math.
- Demonstration schools reduced achievement gaps over a period of 162 school days.

BACKGROUND: Building the skills of tomorrow

DMPS began their partnership with the team of external consultants in 2016-17 to build capacity to prepare today's students and educators for a future in an increasingly global economy and society. DMPS wanted to ensure their schools deliver and sustain high levels of professional and academic performance. Des Moines has large refugee community with over a hundred languages spoken in the school hallways. Their goal was to make rapid progress towards becoming a national model for urban education. During the 2016-17 school year, six DMPS schools became demonstration schools, hereafter called year-two schools. In the 2017-18 school year, 16 additional schools, hereafter called year-one schools, joined the journey.

My work with the consultants this year has really shifted my thinking… . So many times, we underestimate what students can do. The work we're doing with demonstration schools is helping us to confirm that students can do the work.

– District Administrator

SAMPLE: Serving diverse learners

In the 2017-18 school year, approximately 890 teachers in 22 schools participated in the demonstration school initiative, which included professional development training. These sites served 10,431 students. The demographic characteristics of students were 49% female, 51% male, 31% Hispanic, 19% Black, 34% White, 27% English Learners (EL), 17% students with disabilities (SWD), and 82% students eligible for free or reduced-price lunch (FRL).

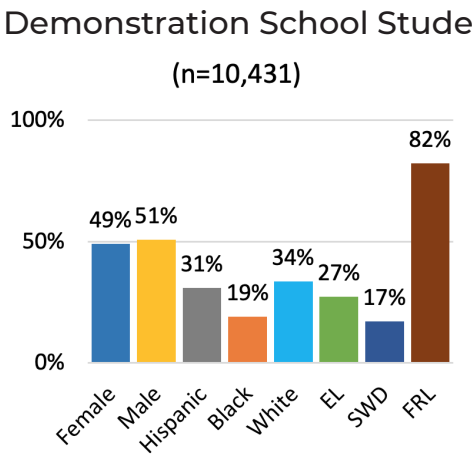

Figure 1. Demographic Characteristics of Demonstration School Students

PROGRAM DESIGN: Building the capacity of leadership teams and teachers

Through the consultants' professional development series, demonstration school leaders and teachers learned techniques to engage in rigorous student-centered teaching in all phases– planning, delivery, reflection, and adjustment. The consultant team assigned school leadership coaches to each school to help support the principal in the implementation of the series. Year-two schools had an consultant-provided leadership coach. Year-one schools used a district employee certified by the consultants as a coach.

The leadership coach consultant focused on coaching the principals and leadership teams at year-two schools and helped train and support district coaches. The consultants guided leadership teams in breaking down standards into smaller targets and ensuring planning time for teachers to create tasks to get students to higher levels of thinking. Teachers also learned how to form and cultivate student academic teams where students challenged each other to use higher-level thinking. Students were guided to own their learning and take responsibility for their progress. By participating in academic teams, students developed the academic, social and emotional learning skills necessary to think critically and independently, as well as work effectively with others.

We have a no-hand-raising rule in our school right now. We also have a new rule: teachers aren't allowed to tell students the answer, they can only ask questions. I see wonderful things happening.
– District Results Coach

RESULTS: Demonstration Schools students outperform by statistically significant margins

What Works Clearinghouse (WWC) Design Standards were used by a certified WWC reviewer to assess the effectiveness of the demonstration schools and professional development series. The evaluation met WWC Design Standards With Reservations. Demonstration school students were matched to like-students in the district served by schools who were not participating in the professional development series and who had similar pretest scores in the Fall. All findings presented are statistically significant ($p<.05$).

Treatment intervention effect sizes were computed. Effect sizes are useful to measure the strength of an intervention on student achievement. It represents the number of standard deviation units by which the intervention group outperforms the control group. They are also used to calculate learning rates and additional learning days.

Learning rates reveal how much more demonstration school students learned compared to the average gain otherwise expected within the same timeframe. Put simply, the learning rate is the treatment effect divided by the average gain otherwise expected. Similar calculations were used to gauge how many extra days of learning were gained and how much demonstration school students closed achievement gaps. Sample sizes (n) in the Figures

only include demonstration school students; however, all demonstration school students were matched to the same number of control students to estimate the program impact. For more details on the methodology and calculations used see Appendix A.

Demonstration school students gained an additional 11 days of learning in reading and six days in math.

Students at all demonstration schools had a statistically significant improvement. Demonstration school students had a 7% improvement in reading and 3% improvement in math over the gain otherwise expected within a period of 162 school days. This improvement translates into an additional 11 days of learning in reading and six days in math for demonstration school students.

Students are working together, they're processing together, and they're really challenging each other. I find myself stepping back and just watching the learning happen.

– 4th Grade Teacher

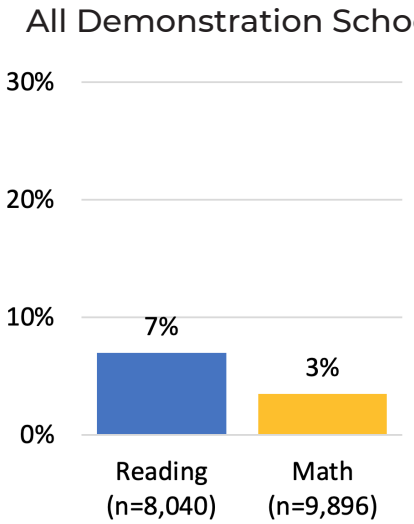

Figure 2. Learning Rates of Students at Demonstration Schools

Year-two Demonstration school students gained an additional 21 days of learning in reading and 10 days in math during the second year of implementation.

Demonstration school students in year-two schools had larger learning gains- a 15% improvement in reading and 6% improvement in math over the gain otherwise expected within a period of 162 school days. This translates into an additional 21 days of learning in reading and 10 days in math. Year-two demonstration schools had an external consultant leadership coach directly coach the principal and leadership teams at these schools. Additionally, students had teachers in their second year of the consultant's professional development.

What we've seen in our most recent trends of academic data is that in our year-two demonstration school performance gaps are actually closing. We met and exceeded our goals for Black males.

– Associate Superintendent

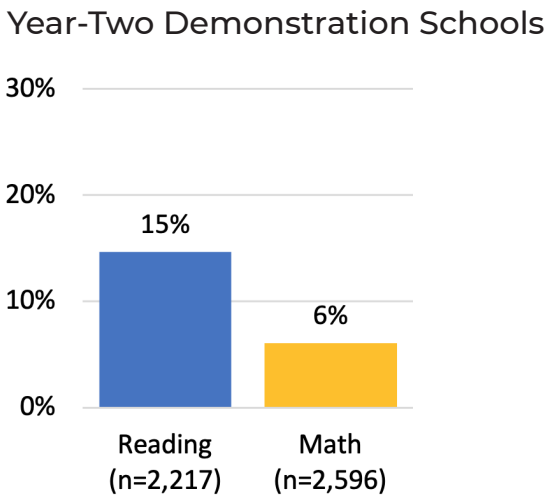

Figure 3. Learning Rates of Students at Year-two Schools

High school demonstration school students gained an additional 61 days of learning in reading and 42 days in math

High school students had the largest learning gains. Two high schools received the treatment-one first-year and one second-year school. High school students in demonstration schools achieved a 37% improvement in reading and 26% improvement in math over the gain otherwise expected within a period of 162 school days. This translates into an additional 61 days of learning in reading and 42 days in math for demonstration school students.

Everyone supports each other; everyone works with each other, everyone helps each other. We leave no child behind.

– High School Teacher

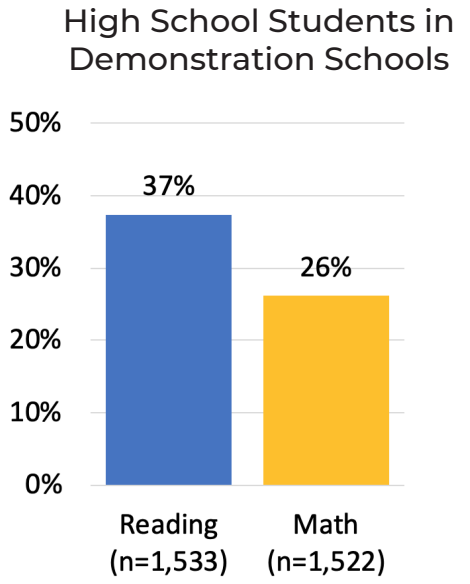

Figure 4. Learning Rates of High School Students

Demonstration schools reduced achievement gaps within 162 school days

- Demonstration schools reduced the Black: White achievement gap by 7% in reading and 6% in math over a period of 162 school days.
- Demonstration schools reduced the Students with Disabilities:Non-Students with Disabilities achievement gap by 6% in reading and 5% in math.
- Demonstration schools reduced the English Learner:Non-English Learner achievement gap by 6% in reading and 4% in math.

Demonstration Schools foster inclusiveness—instead of EL or special education students being pulled out into isolation, these students work in their academic teams within mainstream classrooms.

– Associate Superintendent

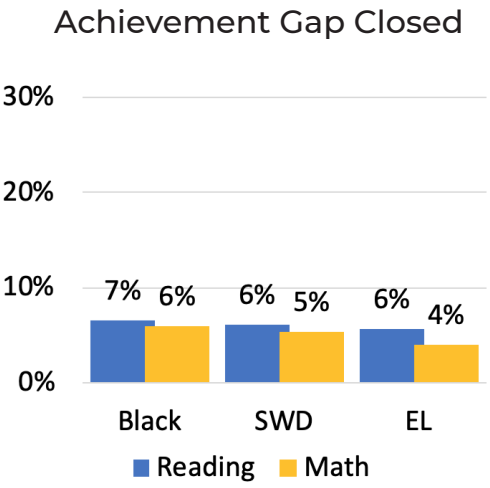

Figure 5. Percent of Achievement Gaps Closed by Student Groups

CONCLUSION: DMPS improved performance by closing achievement gaps

Des Moines Public Schools (DMPS) began their partnership with a team of external consultants in 2016-17 to build capacity to prepare educators and students for a future where critical thinking, social, and emotional skills are essential. The results of this analysis show that DMPS is forging the way to rigor and quickly closing achievement gaps for all students.

While this study is essential in demonstrating their success, the real victory is observing how students learn in their classroom. We invite you to watch the learning unfold in this video of Des Moines Public Schools.

Appendix and References

Appendix: Methodology and Calculations

Quasi-Experimental Design Matching Procedures

Since random assignment was not possible, propensity score matching was used to create a comparable control group by matching demonstration school (treatment) students to like-students in the district who did not receive the treatment (Thoemmes, 2012). While propensity score matching has its limitations, it is the best method available to match like treatment students to a control in the absence of a true control group. Treatment students were matched, using an exact match on grade level and nearest neighbor matching on scale scores for the Fall reading and math NWEA assessments (tolerance < .20). Other covariates in the model included gender, race, ethnicity, students with disabilities (SWD), English learner (EL), free and reduced-price lunch (FRL), and gifted status. This study focuses on achievement specific to the 2017-18 school year as this was the first-year implementation of the professional development series.

Out of 8,270 treatment students who took both the Fall and Spring reading assessments, 97% were matched to a similar control student. Out of 10,060 treatment students who took both the Fall and Spring math assessments, 98% were matched to a similar control student. WWC requires that Quasi-Experimental Designs demonstrate equivalence of the analytical intervention and comparison groups to Meet WWC Group Design Standards With Reservations. WWC standard is to use Hedges' g formula to compute effect sizes. It is defined as the difference between the mean outcome for the intervention group and the mean outcome for the comparison group, divided by the pooled within-group standard deviation of the outcome measure (WWC Procedures Handbook, 2017, p. 13). Hedges' g at baseline indicated equivalence of groups ($g<.25$) on the Fall scale score in reading ($g=.033$) and in math ($g=.024$). To provide more precise estimates of the effects, the Fall scale scores were included in all models to remove any pre-intervention difference between the groups. The outcome of interest used in the analysis was the NWEA Spring scale score.

Students within treatment schools were assigned to the invention as a group; thus, data for analysis are based on the individuals within clusters. Hierarchal Linear Modeling (HLM) was used to adjust for clustering of students within schools (Raudenbush and Bryk, 2002). Covariates in the model included: Fall scale score, gender, race, ethnicity, students with disabilities (SWD), English learner (EL), gifted, and free and reduced-priced lunch (FRL) status. Additionally, one school-level characteristic was included in the models to control for poverty. Specifically, a Distressed Communities Index was incorporated as a measure of poverty, and it includes a factor of seven different indices of poverty combined across five years (The Economic Innovation Group, 2017). Multicollinearity was investigated, and all correlation coefficients were less than .38. Accordingly, there were not abnormally high intercorrelations among the independent variables which could erroneously weaken or strengthen the statistical power of the models.

Estimating the Magnitude of Effects

Hedges' g was also calculated to assess the magnitude of the intervention effects estimated from the HLM analyses. The equation used was slightly different than the Hedges' equation noted for baseline equivalence as it incorporates the adjusted group mean difference from cluster analysis divided by the unadjusted pooled within-group standard deviations of the post-test (WWC Procedures Handbook, 2017, p. E-9). All treatment model effect sizes are noted below and 'ns' indicates values that were not significant.

	Hedges' g	
	Reading	**Math**
All	.030	.022
Year-Two	.063	.038
HS	.054	.046
ES	ns	.031
MS	.039	ns
Black	.046	.046
SWD	.059	.049
EL	.053	.035

Table 1. Hedges' g of Treatment Effects

While there is no inherent substantive meaning to standard deviation units for policymakers, the most common practice is to rely on Cohen's suggestion that effect sizes of about .20, .50, and .80 standard deviation be considered small, medium, and large. As Bloom and colleagues point out (2008, p. 5-6), these guidelines were not relevant to intervention effects in education with Cohen (1988, p. 25) even stating that the suggestions should only be used when there is no better basis for estimating the magnitude of impact. Using Bloom and colleagues (2008) as a guideline, this study creates benchmarks using local norms which can provide more meaningful interpretations of the impact of the program.

Learning Rate Calculations

To create local norms, average student achievement gains were calculated for each grade level using all students in the district who were tested in both Fall and Spring from 1st to 10th grades. Table 2 outlines the average annual gains for reading and math in DMPS from Fall to Spring in 2017-18 school year. Weighted averages are displayed in the last row. Annual gains from the year prior to the treatment would have been optimal; however, not all grade levels were tested in the 2016-17 school year. Consequently, the learning rates calculated in this study may be underestimated because the calculations incorporate a large portion of treatment students.

To calculate the rate of learning in reading for year-two treatment students, an effect size of .063 divided by the average annual gain (.426) equates to a 15% learning rate within 162 school days. School days were calculated by taking the first administration of the pretest minus the last day of administration of the post-test and then counting the number of school days within those dates (no weekends). The number of school days gained was calculated by using the ratio of effect sizes compared to the proportion of days between the pretest and post-test.

Grade	Reading	Math
1	1.281	1.333
2	.938	.977
3	.611	.913
4	.482	.763
5	.408	.574
6	.397	472
7	.243	.374
8	.231	.298
9	.179	.192
10	.112	.162
Wt. Avg.	**.426**	**.626**

Table 2. Average Fall to Spring Gains

Student Group Gap Calculations

Due to the large sample sizes available for Black students, SWD and EL treatment groups, treatment effect sizes for the selected groups were calculated using the same procedures outlined in Quasi-Experimental Design Matching Procedures section. More specifically, each Black student in the treatment group was matched to a Black student in the control group in the same grade level and who had similar Fall scale scores using all other covariates in the model. To satisfy WWC baseline equivalence requirements, the pretest was included in all models to adjust for any differences in groups at baseline. Table 3 shows the matching results with corresponding baseline Hedges' g for the analytic samples.

Out of 1,599 Black students in the treatment groups who took both the Fall and Spring reading assessments, 94% were matched to similar Black students in the control groups. Out of 1,948 Black students in the treatment groups who took both the Fall and Spring math assessments, 97% were matched to similar Black students in the control groups. Hedges' g at baseline indicated equivalence of the treatment and control groups on the Fall scale score in reading (g=.033) and in math (g=.010). Propensity score matching was not used for the EL analyses, as the analytic sample satisfies the baseline equivalence requirement without matching if the models incorporate a statistical adjustment, and they did. Treatment effect sizes for each student group were then calculated as noted in Estimating the Magnitude of Effects section.

Gap benchmarks for student groups were calculated by taking the mean difference in the Fall to Spring achievement for each group/counterpart divided by the standard deviation for each grade level for all tested students (Bloom et

	Reading			Math		
	Treatment Students with Pre/Post	Percent Matched or Included	Baseline Hedges' g	Treatment Students with Pre/Post	Percent Matched or Included	Baseline Hedges' g
Black	1,599	94%	.033	1,948	97%	.010
SWD	1,573	93%	.052	1,734	93%	.027
EL	2,199	100%	.013	2,800	100%	.065

Table 3. Results of Matching Procedures with Hedges' g

al., 2008, p. 20). Table 4 shows the average Fall to Spring Student Group Performance Gaps for the 2017-18 school year. To obtain how much Black students in treatment groups closed the achievement gap, an effect size of .046 for Black students in treatment groups divided by the average Black:White gap (-.701) equates to a 7% closure in the Black:White reading gap within 162 school days. All other performance rates were calculated in the same manner.

	Reading			Math		
Grade	Black:White Gap	SWD:No-SWD Gap	EL:Non-EL Gap	Black:White Gap	SWD:No-SWD Gap	EL:Non-EL Gap
1	-.501	-.736	-.441	-.725	-.627	-.580
2	-.618	-.904	-.664	-.721	-.842	-.640
3	-.721	-.978	-.818	-.804	-.891	-.769
4	-.664	-1.083	-.849	-.717	-1.021	-.799
5	-.738	-1.036	-.985	-.728	-.958	-.884
6	-.739	-1.037	-.907	-.796	-1.082	-.863
7	-.622	-1.091	-.995	-.760	-1.133	-.947
8	-.669	-.949	-1.105	-.756	-1.019	-1.053
9	-.758	-.927	-1.261	-.834	-.963	-1.207
10	-.820	-.675	-1.321	-.890	-.754	-1.205
Wt. Avg.	**-.701**	**-.968**	**-.935**	**-.770**	**-.930**	**-.881**

Table 4. Average Fall to Spring Student Group Performance Gaps

References

Bloom, H.S., Hill, Black, A.R., & Lipsey, M.W. (2008). *Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions.* MDRC. Available at https://www.mdrc.org/sites/default/files/full_473.pdf. Accessed September 1, 2018.

Cohen, J. (1988). *Statistical Power Analysis For The Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications And Data Analysis Methods.* Thousand Oaks, CA: Sage.

The Economic Innovation Group (2017). *The 2017 Distressed Communities Index.*[1] Available at https://eig.org/dci. Accessed September 1, 2018.

Thoemmes, F. (2012). *Propensity score matching in SPSS.* Available at https://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf. Accessed September 1, 2018.

What Works Clearinghouse™ Procedures Handbook (2017). Version 4.0. Available at https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf. Accessed September 1, 2018.

Participating Team Members

*Instructional Empowerment's team members served Des Moines Public Schools (DMPS) under another company led by CEO and Executive Director of Research, Michael Toth, before Instructional Empowerment was founded. Michael Toth and the Instructional Empowerment team have perfected and applied their own evidence-based school improvement approach. Team members who served DMPS include: Year 1 - Kathy Houpt, School Leadership Coach and the following team of shared Faculty Coaches and Staff Developers: Gwen Bryant, Betsy Carter, Catherine Menard, Carla Moore, Deana Senn, Diane Hampel, Jan Matthews, Judy Schmitz, Janis Andrews, Joan Pinkerton, Kathy Taber, Tracy Bollinger, Pat Denholm; Year 2 - Practice Leads: Betsy Carter and Jennifer Reeves; School Leadership Coaches: Janis Andrews and Tracy Bollinger, and the following shared team of Faculty Coaches and Staff Developers: Catherine Menard, Deana Senn, Diane Stultz, Frances Miller, Gwen Bryant, Heidi Trexler, Jan Matthews, Joanna Sozio, Joy Morris, Kara Bentley, Karen Doty, Kelly Harmon, Kendra Strange, Kimberly Wood, Kisha Bellande-Francis, Kristin DeJong, Lisa Lienemann, Lorie Spadafora, Lynn Woods, Michelle Fitzsimmons, Phillip Carr, Ria Schmidt, Roxanne Lawson, Sara Croll, Shannon Pretorius, Susan Schilsky, Terrie Mitev, Theresa Staley, and Tracy Duval.

[1]The findings expressed in this publication are solely those of Dr. Lindsay Devers Basileo and not necessarily those of the Economic Innovation Group (EIG).