Data Analysis and Evaluation of McGraw-Hill's Everyday Mathematics Program and its Impact on M-Step Scores in Michigan Schools

David Marshall, Ph.D., Auburn University

Daniel Henry, Ph.D., Director, Auburn Center for Evaluation

proficient scores that were approximately 5 points higher than their matched peers in both grades.

4. Fourth and fifth grade economically disadvantaged McGraw-Hill students had proficiency rates that were 20% greater than their matched peers.

Limitations of the Evaluation

1. Data for the evaluation were completely dependent on the quality and quantity of information collected and reported by the Michigan Department of Education and MEAP staff.

2. Transience or mobility of the student population in participating schools is a concern. Comparisons made in the evaluation are based on the assumption that children in participating schools received the "treatment" of the *Everyday Mathematics* program, but a more extensive per-pupil analysis of student exposure to *Everyday Mathematics* and student mobility is necessary to fully understand this factor.

3. Intermediate and long-term shifts in knowledge, attitudes, perceptions, and achievements in mathematics may not have resulted from the *Everyday Mathematics* alone. Many schools in Michigan have multiple federal and state initiatives in effect at the same time. Since many of these initiatives are aimed at providing increased academic achievement, this may not have had an impact on student test scores.

4. Interpretations based on statistical significance alone should be made with caution.

Data and Method

This evaluation sought to answer the single evaluation question: *To what extent is the utilization of McGraw-Hill's*

district assessments and other tools, offer a comprehensive view of student progress and achievement. (n.p.)

All data analyses were conducted using Stata version 15. District-level demographic data

districts. This approach also successfully matched 11 of the 12 districts. Two-way ANOVAs were conducted to compare the two variables of interest. The tests found the groups to be statistically similar. This time when the same assumptions were checked, no violations were found. District enrollment was a third variable that was initially included in both matching approaches; however, fewer than half of the treatment districts were matched when it was included. Enrollment was accounted for later in the analysis by simulating individual data. Each of the treatment districts has employed the McGraw-Hill curriculum for a minimum of four years, and most for greater than five years. As such, length of utilization was not considered in these analyses. Taylor School District was the lone district that was not matched in the process. See Table 1 for a description of this district. Table 2 provides a list of districts in each group and Table 3 compares the treatment and matched districts along the two matched variables.

Table 1
Demographics for Taylor School District

Total Enrollment	6320
Non-White	46.6%ii-4 (t)-4 (-)Tj 0.002 Tc -07.2 3 0 Td ( )Tj 37 0 Td (46.6%)Tj 2.58 5 >

Matched

in the treatment and matched districts were compared by conducting a one-way ANOVA. One-way ANOVAs were also conducted for each of five subgroups of third-grade students: (1) students receiving special education services; (2) economically disadvantaged; (3) non-White students; (4) male students; and (5) female students. See Figure 1 for a graph comparing each of the six total groups that were analyzed for third-grade students.

First, all third-grade students across the two groups were compared in terms of whether they earned proficient scores on the MSTEP. No statistically significant differences were found between the scores of students from the treatment districts (M = 58.2%, SD = .49) and the matched districts (M = 57.2%, SD = .50). Two of the subgroup analyses yielded significant findings. Because districts in the treatment and matched groups were already matched on the proportion of economically disadvantaged students and students receiving special education services, two of the subgroup analyses sought to learn if economically disadvantaged and special education students perform differently on the MSTEP in districts that are similar as a whole in terms of these two subgroup populations. However, statistically significant differences existed for students receiving special education services. It should be noted that students in the treatment group (M = 33.7%, SD = .47) performed statistically worse on the MSTEP than did students in the matched group (M = 38.9%, SD = .49). In contrast, White students utilizing the McGraw-Hill curriculum (M = 50.8%, SD = .50) performed substantially better on the MSTEP than students in the matched group (M = 33.5%, SD = .47). No statistically significant gender differences existed between groups. See Table 4 for means, standard deviations, t values, and effect sizes where applicable for each of the analyses.

Figure 1 – Third-Grade M-STEP Proficiency

Table 4.
Means, Standard Deviations, and p-values for 3rd Grade M-STEP Proficiency

	M	SD	*p*-value	*d*
All Students				
Treatment	58.20%	.493	0.47	n/a
Matched	57.15%	.495		
Special Education				
Treatment	31.13%	.463	0.25	n/a
Matched	36.17%	.483		
Economic Disadvantage				
Treatment	33.72%	.473	0.03*	0.11
Matched	38.92%	.488		
Non-White				
Treatment	50.77%	.500	0.00**	0.36
Matched	33.54%	.473		
Male				
Treatment	59.22%	.491	0.43	n/a
Matched	57.66%	.494		
Female				
Treatment	57.21%	.494	0.13	n/a
Matched	54.02%	.499		

*Note: ** p < .01; * p < .05; Alpha significance level set a priori at .05.*

All third llc p.4c6 -c3 y16 0 Tc

Table 5.
Means, Standard Deviations, and p-values for 3rd Grade M-STEP Scaled Scores

	M	SD	*p*-value	*d*
All Students				
Treatment	1304.22	10.72	0.03*	0.06
Matched	1303.51	12.64		
Special Education				
Treatment	1283.92	7.24	0.00**	0.22
Matched	1286.23	17.77		
Economic Disadvantage				
Treatment	1288.20	7.87	0.00**	0.15
Matched	1289.46	7.97		
Non-White				
Treatment	1300.86	14.72	0.00**	0.59
Matched	1292.79	12.41		
Male				
Treatment	1305.12	11.15	0.75	n/a
Matched	1304.98	12.50		
Female				
Treatment	1303.27	10.38	0.00**	0.14
Matched	1301.59	13.17		

*Note: ** p < .01; * p < .05; Alpha significance level set a priori at .05.*

Fourth-grade curriculum

According to McGraw-Hill's *Everyday Mathematics* website (2019), the fourth-grade curriculum focuses on procedures, concepts, and applications in three critical areas:

- Understanding and fluency with multi-digit multiplication and understanding of dividing to find quotients with multi-digit dividends.
- Understanding of fraction equivalence, addition and subtraction of fractions with like denominators, and multiplication of fractions by whole numbers.
- Understanding that geometric figures can be analyzed and classified based on their properties.

The same analyses were conducted for the fourth-grade students comparing students in the treatment district group and matched district group. One-way ANOVAs were conducted for each of the analyses to explore whether differences existed between the two groups. Students who used the McGraw-Hill curriculum (M = 60.85%, SD = .488) outperformed their peers in the matched group (M = 54.89, SD = .498) in terms of scoring proficient on the STEP. All but one of the five subgroup analyses yielded significant findings. See Figure 2 for a graph comparing each of the six groups that were analyzed for fourth-grade students. No differences existed between treatment and matched groups for students who received special education services. For each of the other four subgroups, students utilizing the McGraw-Hill curriculum outperformed their peers in the matched group. Both male and female students separately had higher proficiency rates on the STEP than students in the matched districts, as did economically disadvantaged students. The difference was most substantial for non-White

students; those in the treatment group (M = 49.5%, SD = .500) were proficient on the test more often than their matched group peers (M = 31.0%, SD = .463), and this yielded a medium effect size (d = 0.38). Effect sizes are a standardized measure of practical significance, expressed in terms of standard deviations, or average distance from the mean. Whereas fewer than one-third of fourth-grade matched district non-White students earned a proficient score on the STEP, almost half of similar students did in the treatment group. See Table 6 for means, standard deviations, p-values, and effect sizes where applicable.

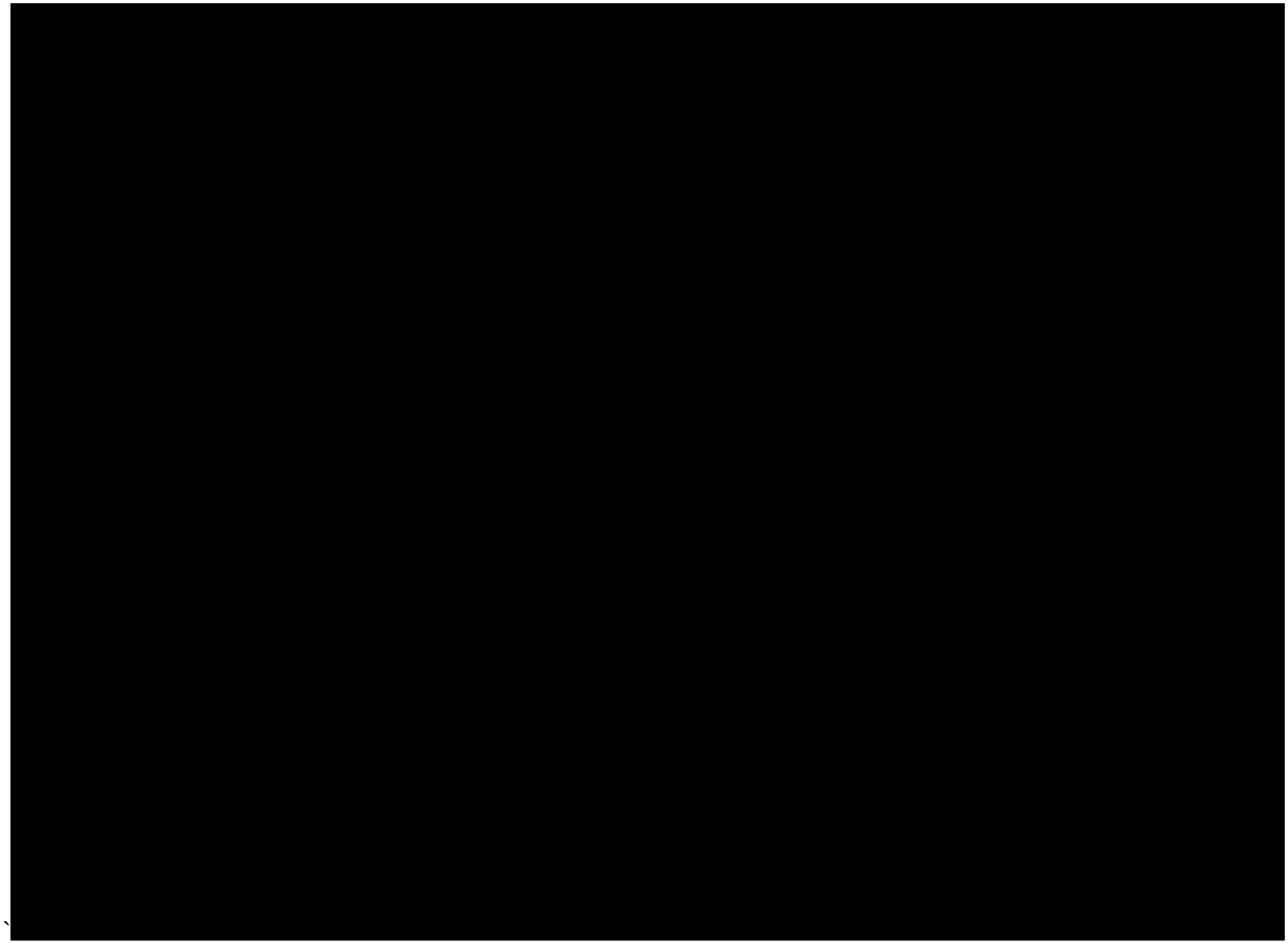

Figure 2 – Fourth-Grade M-STEP Proficiency

Table 6.
Means, Standard Deviations, and p-values for 4th Grade M-STEP Proficiency

	M	SD	*p*-value	*d*
All Students				
Treatment	60.85%	.488	0.00**	0.12
Matched	54.89%	.498		

size (d = 0.52). See Table 7 for means, standard deviations, p-values, and effect sizes where applicable for scaled scores.

Table 7.
Means, Standard Deviations, and p-values for 4th Grade M-STEP Scaled Scores

	M	SD	*p*-value	*d*
All Students				
Treatment	1404.13	10.59	0.00**	0.25
Matched	1401.44	10.62		
Special Education				
Treatment	1381.84	8.29	0.23	n/a
Matched	1380.82	10.48		
Economic Disadvantage				
Treatment	1389.32	8.93	0.00**	0.21
Matched	1387.73	6.04		
Non-White				
Treatment	1399.33	15.10	0.00**	0.52
Matched	1392.30	11.59		
Male				
Treatment	1405.35	11.37	0.00**	0.23
Matched	1402.70	11.85		
Female				
Treatment	1402.94	9.93	0.00**	0.31
Matched	1399.83	10.05		

Fifth

yielded medium effect sizes for the treatment groups. The largest effect size ($d = .41$) was found for non-White students in treatment districts ($M = 43.10\%$, $SD = .495$), who earned proficient M-STEP scores at almost twice the rate of non-White students in matched comparison districts ($M = 24.22\%$, $SD = .429$). See Table 8 for means, standard deviations, p-values, and effect sizes where applicable.

Table 8.
Means, Standard Deviations, and p-values for 5th Grade M-STEP Proficiency

	M	SD	*p*-value	*d*
All Students				
Treatment	55.14%	.497	0.00**	0.13
Matched	48.67%	.500		
Special Education				
Treatment	18.53%	.389	0.30	n/a
Matched	22.30%	.418		
Economic Disadvantage				
Treatment	28.27%	.450	0.02*	0.12
Matched	23.38%	.424		
Non-White				
Treatment	43.11%	.495	0.00**	0.41
Matched	24.23%	.429		
Male				
Treatment	56.97%	.495	0.00**	0.21
Matched	46.61%	.499		
Female				
Treatment	51.29%	.499	0.01**	0.11
Matched	45.77%	.499		

*Note: ** p < .01; * p < .05; Alpha significance level set a priori at .05.*

Fifth-grade data were also analyzed in terms of whether differences existed between treatment and matched districts on mean M-STEP scaled scores. Overall, students enrolled in treatment districts (M = 1498.52, SD = 11.36) had significantly higher mean scaled scores than their peers in matched comparison districts (M = 1495.66, SD = 10.92), representing a medium effect size (d = 0.26). When subgroups were analyzed, no significant differences were found between treatment and matched districts for students receiving special education services or economically disadvantaged students. However, gender differences were found between the two groups of school districts. Both male and female students in treatment districts outscored their counterparts on the M-STEP, and both analyses yielded medium effect sizes. Similar to the findings from analyses conducted for third grade and fourth grade students, the largest difference between treatment and matched districts was found for White treatment district students, for

attention that has been given to gender differences in STEM. A large body of literature (e.g. Wong & Degol, 2017) has found females less likely to pursue careers in STEM; as such, curricula that support female learning in STEM subject areas, including math, are worth

References