education northwest

CREATING STRONG SCHOOLS & COMMUNITIES

Oregon MESA: Improving Grades in Science and Mathematics

By Jason Greenberg Motamedi and Malkeet Singh, Education Northwest
October 31, 2016

Oregon Mathematics, Engineering, Science Achievement (MESA) focuses its afterschool programs with middle and high school students on inventions that address key problems in developing countries, such as sustainable lighting, water transportation, water filtration, and prosthetics. To this end, MESA created four curricular units in which students collaborate to design, build, evaluate, and present a usable product designed for a particular client. For example, some students, working on the prosthetic limb curriculum, designed and built prosthetic arms for children who had lost limbs due to land mines.

With the support of a grant from the Oregon Community Foundation, MESA expanded its afterschool program to four Salem-Keizer Public Schools middle schools in the 2014/15 and 2015/16 school years. The goal of this expansion was to help underrepresented minority and low income students achieve scholastic success, leadership skills, and social support for college enrollment through (1) afterschool STEM (science, technology, engineering and math) programs led by teachers from the school served; (2) science and technology competitions; (3) family involvement and advocacy; and (4) mentorship by local college students.

At the request of MESA, researchers at Education Northwest conducted an evaluation of the impact of MESA's afterschool program on the educational outcomes of MESA participants in Salem-Keizer middle schools (referred to as "MESA students"). To do this, we collected a student-level dataset from Salem-Keizer and examined how participation in MESA affected student outcomes. Specifically, we used statistical methods to identify a comparable group of students with similar characteristics but who did not participate in the MESA program. We then compared these students' school attendance, test performance, grades, and behavior to those of MESA students.

Key Finding: MESA students had higher grades in science and mathematics classes than their peers. We did not detect any other statistically significant differences between MESA students and a matched sample of other students.

WHAT DID WE DO?

We compared MESA students to a "matched sample" of Salem-Keizer students who did not participate in MESA's afterschool program but were very similar in other ways to determine the impact that MESA had on participants' school attendance, test scores, grades, and behavior.

What is a matched sample?

A matched sample is a way of creating a rigorous and fair—apples to apples—comparison to determine the impact of a program, such as MESA, on participants. Matching ensures that we are comparing students who had similar demographic characteristics, as well as similar test scores and grades, before they participated in MESA.

The matching method we chose uses a "propensity score" to measure the similarity between students, and, therefore, their likelihood of joining MESA. This is called a "quasi-experimental design," because instead of an experimental design—which randomly selects which students participate in MESA and which do not—we used statistical methods to identify students who would likely have joined.

However, unlike random assignment, a matched sample cannot guarantee that all characteristics that could affect student performance are accounted for. While we can test to see if the matched sample of students and MESA students are similar before joining MESA, we cannot do this with characteristics that are not measurable from district datasets, such as student motivation or parent engagement.

This means that the limitation of using matched samples is that we cannot completely attribute differences in student outcomes to the program itself. In other words, **we are certain that the students who participated in MESA received higher grades in science and mathematics than their peers, but we cannot be certain that participation in MESA was the cause of these improved grades;** there may have been other causes that we did not account for.

WHY DID WE USE MATCHING?

Random assignment is the most rigorous method of determining the program impact. Random assignment of enough students ensures that all characteristics that could affect student performance are balanced and equal between students who participate in the program and those who do not participate.

However, random assignment is not always practical or ethical. In its place we used statistical methods to create a matched sample of students who are similar to the students in the program in all observable ways, except that they did not join the program.

How did we match students?

We used eight "covariates" (student-level data points that are possibly predictive of student learning and behavior) from 2013/14 school year—one year before MESA was offered in Salem-Keizer schools—to ensure that the students were well matched:

- English language arts state test scores
- Mathematics state test scores
- Federal race/ethnicity category
- Gender
- Eligibility for the federal lunch program
- Eligibility for special education
- English learner status
- Number of suspensions

With these characteristics, we matched each of the 89 MESA middle school students, who participated in MESA's afterschool program during the 2014/15 or 2015/16 school years and had complete grades and test scores from the 2013/14 school year,[1] to their "nearest neighbors," who were as similar as possible (Table 1). To increase our ability to recognize statistically significant differences ("statistical power"), we combined all MESA students into one group, regardless of whether they participated in MESA during 2014/15, 2015/16, or in both years.[2]

Table 1. Baseline results in 2013/14 were well matched

	Salem-Keizer (N = 6,792)	*Matched Sample (N = 89)*	*MESA Students (N = 89)*
Average state test scores (OAKS)			
English language arts	222	223	223
Mathematics	224	226	226
Percent of students who are			
Underrepresented	46%	75%	75%
Female	51%	32%	38%
Federal lunch program	58%	65%	74%
Special education	14%	16%	12%
English learners	16%	18%	21%
Average number of suspensions	.08	.08	.08

Note: We combined American Indian, Black, Latino, and Pacific Islander students into the category "Underrepresented" because of the small sample size, to protect student anonymity and increase our ability to detect differences.

We examined the similarity of the matched sample to MESA students, and found that the differences were close to zero, which means that the two groups were very well matched. Differences in grades and absences from 2013/14 between the matched sample and MESA students were also close to zero.

How did we check the accuracy of the matching?

We checked that our matched sample of students was appropriate and our findings were accurate by using a second matching method. This method also uses "propensity scores." However, rather than choosing real students as a "nearest neighbor" comparison, this method adjusts the scores, grades, and other characteristics of both MESA students and the matched sample, and then compares these adjusted scores for all students in our sample. The Appendix, which follows this narrative, provides a technical summary of this method and findings.

WHAT DID WE FIND?

We found that MESA students had *higher grades in science and mathematics classes than their peers in the 2015/16 school year,* as well as a higher grade point average in their core English, mathematics, science, and social studies classes. There were *no other statistically significant differences* in test scores, attendance, or suspensions between MESA students and their matched peers (Table 2).

Table 2. MESA students had higher grades in mathematics and science in 2015/16 then their peers

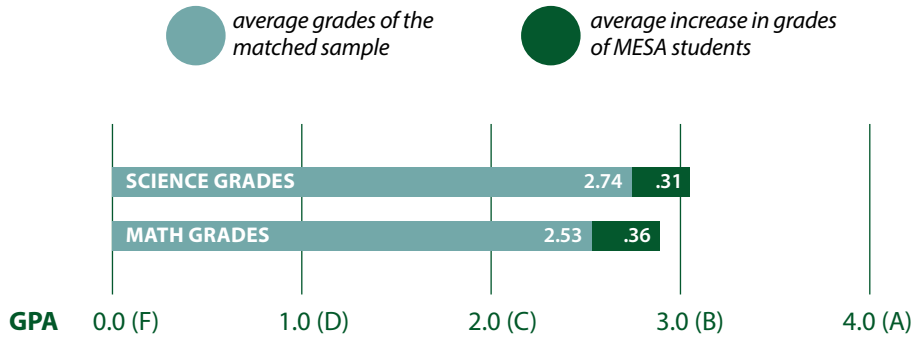
	Salem-Keizer (N = 6,792)	*Matched Sample (N = 89)*	*MESA Students (N = 89)*
Average state test scores (SBAC)			
English language arts	2563	2560	2559
Mathematics	2545	2556	2552
Average number of			
Suspensions	0.08	0.08	0.08
Grade point average			
English	2.89	2.67	2.82
Mathematics*	2.74	2.53	2.89
Science*	2.87	2.74	3.05
Core subjects*	3.07	2.91	3.13

* Statistically significant difference between MESA and matched sample ($p \leq .05$)

Even after accounting for demographic differences, grade level, and prior academic achievement, the association between MESA participation and grades in mathematics ($p = 0.01$) and science ($p = 0.03$) were statistically significant. MESA students' overall grade point average in core classes were also significantly higher ($p = 0.02$) than their peers.

MESA students' grades in science and mathematics were about 11 percent (one third of a grade) higher than their peers. In other words, after participation in MESA, an average student's grades moved from a C+ to a B- in science and from a C to a C+ in mathematics (Figure 1).

Figure 1. MESA students had higher grades in science and mathematics classes than their peers



Were students affected differently?

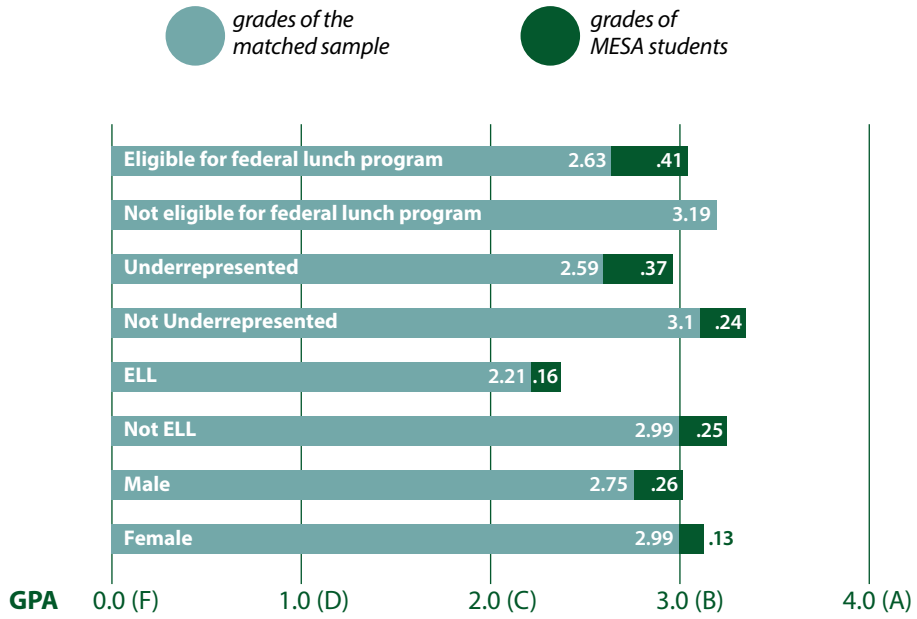
Among MESA participants, we did not find any statistically significant differences in how MESA affected grades by students' gender, race/ethnicity, eligibility for the federal lunch program, or English learner status. This means that MESA is likely to be effective regardless of student characteristics.

Moreover, MESA students consistently outperformed their peers across almost all demographic groups in science (Figure 2) and mathematics (Figure 3). In other words, the effect of MESA was consistent across almost all groups. However, we cannot rule out the possibility that MESA had different effects on different groups of students. This is because demographic analysis usually requires a large sample size, and we likely did not have the number of MESA students required to detect a statistically significant effect.

While the differences among student groups were not statistically significant, they were observable. For example, among students eligible for the federal free lunch program (students affected by poverty), MESA students outperformed their peers in both science and math by more than a third of a grade (.41 in science and .35 in mathematics). While this did not raise their grades to the level of students not eligible for the federal free lunch program, it brought them closer and ameliorated some of the differences.

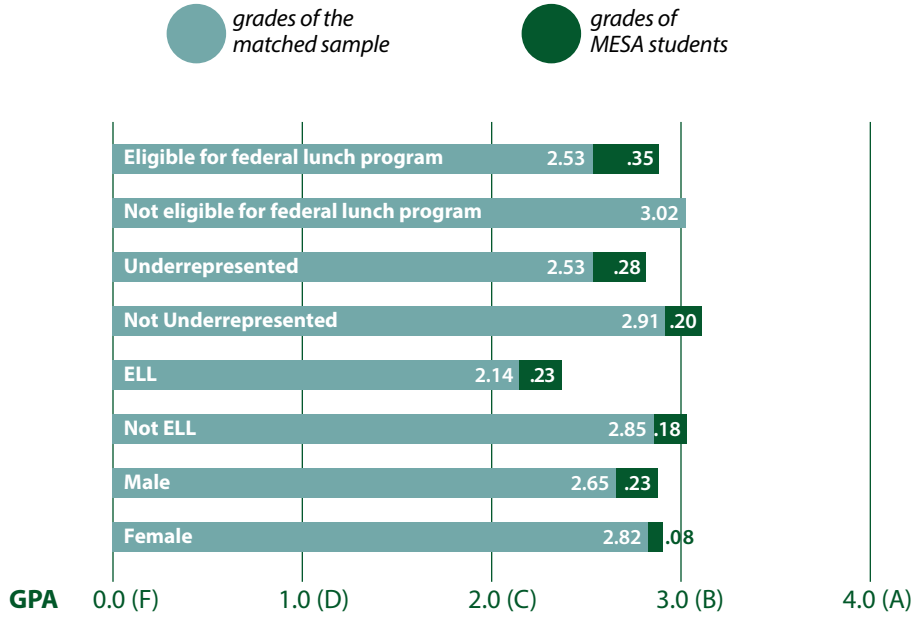
MESA is likely to be effective regardless of student characteristics.

Figure 2. MESA students had higher grades in science in almost all demographic groups



Note: We combined American Indian, Black, Latino, and Pacific Islander students into the category "Underrepresented" because of the small sample size, to protect student anonymity and increase our ability to detect differences.

Figure 3. MESA students had higher grades in mathematics in almost all demographic groups



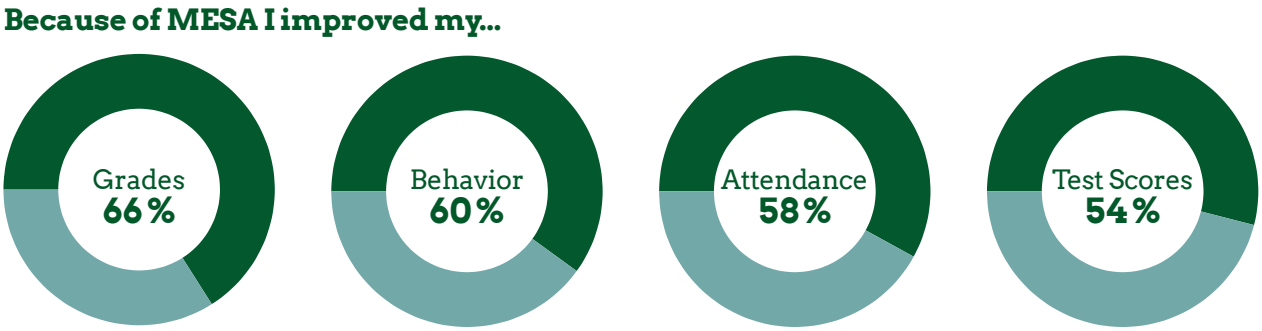
Note: We combined American Indian, Black, Latino, and Pacific Islander students into the category "Underrepresented" because of the small sample size, to protect student anonymity and increase our ability to detect differences.

Do these findings agree with previous evaluation results?

This is the first impact evaluation of MESA. However, Education Northwest has been partnering with MESA for over eight years, collecting data on student perceptions of the effect of MESA on their attitudes, skills, and performance in school. The findings from this impact evaluation appear to corroborate survey findings among MESA students.

In our June 2015 and 2016 surveys, two thirds of MESA students (66%) in Salem-Keizer schools agreed that participation in MESA improved their grades. Slightly fewer reported that it improved other school outcomes (Figure 4). These data do appear to substantiate our findings. However, student responses were less positive on questions around impact than on questions about attitudes and skill growth. For example, more than 80 percent of students agreed or strongly agreed that their confidence in using one or more of their invention skills increased because of MESA.

Figure 4. Two thirds of MESA students agreed that MESA improved their grades.



Source: Surveys of MESA students in Salem-Keizer Public Schools, 2014/15 and 2015/16

WHAT'S NEXT?

MESA students achieved higher grades than their peers in science and mathematics, and it appears that MESA may be a cause of this difference. Therefore, we believe that MESA should continue expanding its reach, serving more students in its afterschool program, especially those from underrepresented groups and those affected by poverty.

We have four recommendations for a future research agenda for MESA to collect rigorous data that supports MESA's efforts to ensure that all students—especially those who most need support—benefit from participation.

We recommend that MESA consider conducting a randomized controlled trial to measure its impact on student achievement with certainty. The ability to attribute causation with greater certainty can only be determined by randomly assigning students to participate or not participate in MESA. While randomization may not always be practical, there may be advantageous opportunities that would make such a controlled trial possible—such as when there are more students ready to enroll in MESA than there are positions available; or, alternatively, by staggering enrollment over a few years, with some students assigned to an earlier cohort and some to a later cohort.

We recommend that MESA improve and standardize student attendance records. We found that records of MESA attendance were not consistent, and that not all MESA advisors recorded student attendance at all MESA meetings. Therefore, we could not use attendance as a "dosage effect" to determine if more MESA means more improvement. We recommend that MESA work with its advisors to ensure that student attendance is constantly recorded.

We recommend that future MESA studies examine outcomes from larger numbers of students. In the current study, we could not accurately measure the relationship of student characteristics to outcomes because of the small sample size. (89 MESA students had complete records.) For example, it appears that MESA had a larger impact on the grades of students affected by poverty (those who qualify for the federal lunch program) than others; however, we did not have the "statistical power" to test for statistical significance because of our limited sample size. Examining outcomes from a larger number of MESA students would allow us to explore the relationship of student characteristics to outcomes in more detail. MESA could increase numbers by including students from multiple districts in future studies.

We recommend that future MESA studies examine longitudinal outcomes. This study examines the relationship of MESA performance and student outcomes at one point in time—2015/16—the year after participation for most MESA students. Participation in MESA may also have a long-term impact on students' achievement. A longitudinal study that tracks the outcomes of MESA students to high school graduation and beyond to college and career, may reveal the long-term impacts of MESA participation.

TECHNICAL APPENDIX

We conducted three separate analyses to gauge whether our findings were consistent and comparable. The method outlined in the text of this report used the most common matching method—one MESA student ('treatment") matched to a non-MESA student ("control"). A second analysis used a similar method, but matched four non-MESA students to each MESA student. It has been suggested that four matches reduces the mean-squared error, providing more precision to the analysis.[3] The number of matches did not alter the findings of our original results.

We also checked our findings using augmented inverse-probability weighting (AIPW). This method provides an alternative way to estimate MESA's effects on student outcomes by reducing the selection bias inherent in observational studies. The main objective is to make MESA students and the matched sample as similar as possible in terms of the estimated propensity scores.

A post-match analysis of AIPW showed that all covariates were sufficiently balanced between the two groups; standardized differences of the means of the covariates were close to zero (Table 3). Robust standard errors were used in the outcome analysis for testing statistical significance of the post-match outcome analysis, and grade level was controlled for in the outcome analysis as fixed effects.

The results of this method were slightly different than the one-to-one or one-to-four matching. The AIPW found science grades to be statistically significant and positive direction in favor of MESA students, while math grades were positive in favor of MESA students but did not reach statistical significance. Subgroup analysis did not find any interaction effect between the demographics and the outcomes.

Table 3. Baseline and outcome analysis using AIPW

	Non-MESA students weighted and matched (N = 3442.2)		*MESA Students weighted and matched (N = 3349.8)*	
Covariates (2014)	**Mean**	**SD**	**Mean**	**SD**
English language arts OAKS test scores	222	15.27	222	9.89
Mathematics OAKS test scores	224	15.24	225	10.11
Underrepresented	46%	-	48%	-
Female	51%	-	54%	-
Federal lunch program	58%	-	61%	-
Special education	14%	-	16%	-
English learners	16%	-	16%	-
Outcome variables (2016)	**Mean**	**Robust SE**	**Mean**	**Robust SE**
English language arts SBAC test scores	2563	1.24	2554	7.85
Mathematics SBAC test scores	2545	1.47	2547	7.56
English grades	2.89	0.01	2.89	0.08
Mathematics grades	2.74	0.01	2.86	0.09
Science grades*	2.87	0.01	3.16	0.13
Core subjects grades	3.07	0.01	3.09	0.07
Suspensions	0.08	0.01	0.09	0.04

* Statistically significant difference between MESA and matched sample ($p \leq .01$)
Note: We combined American Indian, Black, Latino, and Pacific Islander students into the category "Underrepresented" because of the small sample size, to protect student anonymity and increase our ability to detect differences.

We chose to represent the findings from the one-to-one match in the body of this report because the method is rigorous and the findings are easily understood and displayed. The results from the AIPW are difficult to display, since this method adjusts the scores, grades, and other characteristics of both MESA students and the matched sample, and then compares these adjusted scores.

END NOTES

[1] Not all MESA students had complete records. Approximately one third (31 percent) of MESA students were missing grades and/or test scores. These students were excluded from the analysis.

[2] There were no statistically significant differences in demographics or outcomes between students who participated in MESA in either 2014/15 or 2015/16. Likewise, there were no statistically significant differences in the demographics or outcomes between students who participated in MESA one year compared to two years.

[3] Abadie, A. & Imbens, G. (2002). Simple and bias-corrected matching estimators for average treatment effects. [Technical Working Paper 283]. Cambridge, MA: National Bureau of Economic Research.