RESEARCH**REPORT**

AN EMPIRICAL EDUCATION REPORT

Alabama Science in Motion and Student Science Achievement in High School

A CORRELATIONAL REPORT IN ALABAMA

December 2023

Val Lazarev

Adam Schellinger

Jenna Zacamy

Empirical Education Inc.

Acknowledgements

We are grateful to the staff at the Alabama State Department of Education and the Alabama Science in Motion team for providing the data needed to conduct this research. The research was sponsored by the University of Montevallo, an Alabama Math, Science, Technology Initiative regional in-service education center, which provided Empirical Education Inc. with independence in reporting the results.

About Empirical Education Inc.

Empirical Education Inc. is an education research firm that uses its deep knowledge of education and its expertise in analytics, research design, and engineering, to help K-12 schools improve their programs and processes.

Table of Contents

Reference this report: Lazarev, V., Schellinger, A., & Zacamy, J. (2023). Alabama Science in Motion and Student Science Achievement in High School. A Correlational Report in Alabama. (Empirical Education report number: Empirical_ UM_AMSTI-7040-FR1-2023-O.1). Empirical Education Inc. https://www.empiricaleducation.com/amsti/

Principal Findings

This report leverages state educational agency data to explore the association between Alabama Science in Motion (ASIM) implementation and student outcomes. We conducted this study in Alabama to evaluate the potential of teacher implementation of ASIM lab activities to improve high school level student achievement in science.

- There was a significant positive association between ASIM implementation and student scores on the ACT Science test. The percentile rank of students of fully-implementing ASIM teachers is estimated to be higher by 10 points than the rank of students taking classes with non-implementing teachers.
- The number of students achieving college readiness in science (an ACT Science score of 26 or higher) is estimated to be 58 percent higher if all science classes are taught by fully-implementing ASIM teachers.
- ASIM implementation is positively associated with student outcomes for students identified in all racial/ethnic groups.
- The ACT Science subscale most strongly affected by ASIM implementation is "Evaluation of Models."

Background

Through our partnership with the Alabama State Department of Education (ALSDE) and the Alabama Math, Science, Technology Initiative (AMSTI) team, we have completed several prior studies on the impact of AMSTI in upper-elementary, middle, and high schools. These studies include a multi-year randomized control trial conducted under a federal Department of Education contract that involved 82 Alabama schools and over 700 teachers and showed an overall positive effect (Newman et al., 2012). More recently, we completed a study that investigated (1) the impact of AMSTI on math, reading, and science achievement for students in grades 3 through 8; (2) if the impact on student achievement varied across important subgroups of students; and (3) the impact on early career teachers. That study also showed positive impacts for students and provided several recommendations for improvement and suggestions for follow-on research (Lazarev et al., 2019).

This study continues the series focusing on the Alabama Science in Motion (ASIM) initiative. ASIM is a program that plays a significant role in the professional learning of high school science teachers and hands-on learning for high school science students. Since its inception, ASIM has trained teachers in almost all high schools in the state to perform lab activities in science classrooms that are expected to improve student learning and increase interest in science.

Research Questions

This study addresses the following questions.

1. Is there an association between ASIM implementation and science achievement for students in grade 11?
2. Does the association of ASIM implementation with student achievement vary across student groups?
3. Where can improvement efforts be applied in the ASIM process to produce greater program effectiveness?

Study Description

OUTCOME MEASURES

In this study, we extend our analysis of the impact of ASIM into high schools by focusing on the ACT Science scale scores and subscales, as well as the probability of reaching or exceeding the score of 26, which is known to be an indicator of future success in college in STEM subjects.[1]

STUDY DESIGN

This study follows a correlational design aimed at establishing the statistical association between a single metric of teacher preparation and student outcomes. Unlike experimental and quasi-experimental studies, it does not compare treatment and control groups. It focuses entirely on students exposed to the program (in this case, the implementation of ASIM labs)—and the differences in outcomes among them that can be attributed to the metrics of program implementation—making appropriate adjustments for differences in student and school characteristics and pretest scores.

The results of such a study are used to predict potential outcomes under the best possible conditions. In this study, such conditions are achieved when each science teacher fully implements ASIM. We compare those potential outcomes to imputed outcomes for students in the classes of teachers with no ASIM lab implementation. Positive results should be viewed as showing potential (promise of effectiveness) rather than proving effectiveness, because of the wide variability in program implementation.

DATA COLLECTION AND SAMPLE IDENTIFICATION

This study used existing data from the information system at ALSDE and from the ASIM team.

Data we collected for this study consisted of 94,843 individual records for two consecutive cohorts of students taking the ACT in the spring of 2021 and 2022, roster data for all science classes taken by those two cohorts in grades 9 through 11, and science teacher ASIM lab implementation records. Student records contained student demographics, pretest scores (eighth grade math and reading state test scores), and science outcomes (ACT Science composite scores and three subscale scores). We linked teacher data to student records using roster files. We linked additional school data from the NCES database.

We included a subset of student records—that had available pretest scores, ACT test scores, and science course rosters—in the sample for analysis, representing 68% of the original sample. The reduction in the size of the sample did not change its composition compared to the whole population of Alabama eleventh graders. Parameters of the final analytic sample are in Tables 1–2.

---

[1] What are the ACT College Readiness Benchmarks? https://www.act.org/content/dam/act/unsecured/documents/pdfs/R1670-college-readiness-benchmarks-2017-11.pdf

TABLE 1. SAMPLE SIZES

Category	All data	Analytic sample
School systems	157	145
Schools	418	384
Students	94,843	65,227

TABLE 2. CHARACTERISTICS OF THE ANALYTIC SAMPLE

Category	All data	Analytic sample
Free/reduced-price lunch	41.3	40.3
English language learner	2.2	1.4
Special Education	10.1	8.7
White	58.1	59.1
Black	31.2	30.7
Hispanic	4.1	3.9
Other ethnicities	6.6	6.3

We calculated the main indicator of student exposure to ASIM implementation—average ASIM score—in the following two steps. First, using teacher lab records, we calculated a measure of ASIM implementation—ASIM score—for each teacher in each subject (biology, chemistry, and physics) by dividing the number of ASIM labs they used by the maximum number used. This measure ranges from 0 to 1. Then, using roster data, we calculated the average ASIM score every student by averaging the ASIM scores of all their science teachers (i.e. teachers that taught classes they were enrolled into). Student ASIM scores typically range from 0 to 1. However, in some cases, a student's ASIM score can exceed 1. This can happen when a student took several courses with a teacher who had implemented extensive ASIM lab activities in more than one subject. The maximum student ASIM score in the sample equals 2, but the proportion of students with scores exceeding 1 is very small: less than 0.1%.

Some characteristics of student exposure to ASIM lab implementation are in Table 3. Note that ASIM scores reflect the relative teacher usage of ASIM lab activities but do not convey information on the actual number of ASIM lab activities undertaken in each class.

TABLE 3. SAMPLE SIZES

Implementation	Value
Students who had no ASIM-implementing science teachers, % total	10
Students who had only ASIM-implementing science teachers, % total	38
Average student ASIM score (student level)	0.13

Data Analysis

We analyzed the association between ASIM implementation and student outcomes using a hierarchical mixed-effects linear regression model (generalized linear for achieving college readiness). We used ACT composite and subscale scores and a binary indicator of reaching the score of 26 as outcome variables. Student and school characteristics, pretest scores, and the total number of science courses taken were the covariates, in addition to the ASIM score (described above). The models make adjustments for students' clustering in schools and school systems.

The models estimate the association between student test scores and a one-unit increase in ASIM scores. Since the maximum ASIM score equals one, the estimate of the effect of a one-unit increase in the ASIM score is equivalent to the estimate of the hypothetical effect of full implementation by all teachers (using all labs across all courses). We therefore use these estimates as the estimates of the potential program effect at full implementation. We also estimate the effect of actual implementation (average ASIM score in the sample), comparing the outcomes for an "average" student to the hypothetical outcomes for students whose teachers do not implement ASIM lab activities. The results are presented as percentile gains for a hypothetical student who would score at the 50th percentile if they were enrolled only in a class taught by a non-implementing teacher. These estimates are presented in Table 4 for all students and for student groups. We present the effect of ASIM on the probability of achieving college readiness (an ACT science score of 26 or higher) as the comparative probabilities of success for students taught by non-implementing vs. fully implementing teachers, as well as the projected increase in the number of successful students.

We estimated effects for student groups by including interaction terms in the model. This allowed the identification of differences in the association between ASIM implementation and student characteristics. The effects reported here estimate the differences in outcomes between two 'average students' who only differ in one characteristic (e.g. socioeconomic status) but are otherwise identical.

Results

We found evidence that ASIM implementation, as measured by the ASIM score, is positively associated with student outcomes. The average student ASIM score in the sample is 0.13. This level of implementation is associated with a 0.18 test score gain on the ACT Science test, or one percentile (effect size = 0.03). We estimate that studying with fully-implementing teachers (an average ASIM score of 1) would be associated with a 1.35 test score gain, or 10 percentiles on the ACT Science test (effect size = 0.27). The level of confidence we have in these estimates is high.

ACT SCIENCE SCORES

ASIM implementation is positively associated with student outcomes on the ACT Science test in all student groups. Despite large differences between some complementary groups, most of these differences are not statistically significant due to large standard errors of estimates, with the exception of the differences between students eligible for free- or reduced-price lunch (FRPL) and non-FRPL students, and rural vs. urban/suburban schools. In other cases, we do not have confidence that estimated differences are not a statistical artifact. While the estimated improvements are larger for some student groups, we do have strong confidence that all student groups are positively affected by ASIM.

TABLE 4. STUDENT RESULTS OVERALL AND BY GROUP, ACT SCIENCE TEST

Category	Average effect (percentile)	Projected effect of full implementation (percentile)	Significant differential from complement
All	1	10	-
Female	1	9	No
Male	2	12	No
ELL	2	14	No
Non-ELL	1	10	No
Special education	1	6	No
Not special education	1	11	No
FRPL	1	8	Yes
Non-FRPL	2	12	Yes
Rural/town	1	8	Yes
Urban/Suburban	2	14	Yes
Asian	1	16	No
Black	1	5	No
Hispanic	2	9	No
Other or unknown	2	12	No
White	1	13	No

ACT SCIENCE SUBSCALE SCORES

In addition to the analyses related to the ACT Science score reported above, we performed similar analyses with the three subscales of ACT Science scores. The results presented in Table 5 show that ASIM implementation has strong positive associations with the Evaluation of Models and Interpretation of Data ACT subscales. It also shows a weaker negative association with the Scientific Investigation subscale. Effect sizes are estimated for students taught by fully implementing teachers (an ASIM score of 1).

TABLE 5. ASSOCIATION WITH ACT SUBSCALES

ACT subscale	Effect size	*p* value
Evaluation of Models	.44	<.001
Interpretation of Data	.32	<.001
Scientific Investigation	-.10	.01

ACT STEM BENCHMARK

Finally, we estimated a model that predicts the probability of a student achieving a score of 26 of higher on the ACT (an indicator of future success in college in STEM subjects). We used it to compare estimated probabilities of success for students taught by non-implementing teachers (ASIM score of 0) and fully implementing teachers (ASIM score of 1). Results presented in Table 6 show a significant increase in the chances of success.

TABLE 6. ASIM AND COLLEGE READINESS

Outcome	Estimate
Predicted probability of scoring 26 or higher on ACT Science, non-implementing teachers	6.1%
Predicted probability of scoring 26 or higher on ACT Science, fully implementing teachers	9.6%
Relative increase in the chance of success in collect STEM subjects	58%

Conclusion and Recommendations

Results of this study present strong evidence of promise that ASIM can improve science outcomes across student groups. The results suggest however that the student groups that are traditionally in the greatest need of access to high-quality and engaging science instruction and materials—such as students in rural schools or economically disadvantaged students—may benefit from the program less than students in urban/suburban schools or more affluent (non-FRPL) students.

The study sample is large enough to obtain statistically significant results for student groups. The reliability of this study's results may be negatively affected by a substantial proportion of students excluded from the analysis due to lack of test scores or science course-taking data.

The finding that only two of the three ACT subscale scores are positively affected by ASIM implementation suggests that some aspects of the ASIM lab activities may require additional improvement. In the interpretation of the results of this study, it is important to remember that this is a non-experimental study, with no defined treatment and control groups, and that the reported effects are projections based on correlational results.

References

ACT. (2021). ACT College and Career Readiness. https://www.act.org/content/act/en/college-and-career-readiness.html

Alabama Math, Science, and Technology Initiative (AMSTI). (2021). AMSTI website. https://www.amsti.org/

Lazarev, V., Schellinger, A., Zacamy, J., & Newman, D. (2019). *Efficacy of the Alabama Math, Science, Technology Initiative (AMSTI) on Math, Science, and Reading Achievement. A Report of a Quasi-experiment in Alabama.* (Empirical Education report number: Empirical_ UM_AMSTI-7038-FR1-2019-O.1). San Mateo, CA: Empirical Education Inc. Retrievable from bit.ly/AMSTIreport

Newman, D., Finney, P.B., Bell, S., Turner, H., Jaciw, A.P., Zacamy, J.L., & Feagans Gould, L. (2012). *Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI). (NCEE 2012–4008)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

What Works Clearinghouse. (2018). *Standards handbook (Version 4.0)*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Appendix A. Technical Details

The tables below present detailed results of the analysis of the associations between ASIM scores and ACT scores. The main results are reported as the "effect per unit" of the ASIM score. The maximum ASIM score for a student does not normally exceed 1. Therefore, this estimate is also the estimate of the effect of having all science teachers fully implementing. This effect is on the ACT Science score scale. Thus, the effect of 1.35 for all students means that the average difference in test scores between students of non-implementing and fully implementing teachers is 1.35. This measure is transformed into the effect size by dividing it by the standard deviation of the ACT scores in the analytic sample. Effect sizes can be compared across studies using different outcome measures.

Multiplying the unit effect by the actual average ASIM score in the sample—0.13—yields the estimate of the actual average level of ASIM implementation. The *p* value is the measure of the precision of the results or the strength of evidence that the effect in question is statistically different from zero. Conventional interpretation is that a *p* value of .05 or less signifies strong evidence, and *p* values above .05 but less than .20 provide limited evidence. Higher *p* values imply that our results provide no reliable information about the association of teacher ASIM implementation and outcomes, since the probability that the true effect is zero—or even has an opposite sign—is too high. Higher *p* values (lower precision of the results) are typical when the student group is small. High or low effect estimates should be ignored when the *p* values are greater than .20.

Prior to performing the analysis of association between ASIM scores and student outcomes, we performed the analysis to detect possible selection bias. The presence of such bias resulting from non-random variation in ASIM scores would render the results of the association analysis invalid since it would imply that observed effects may be due to differences in school characteristics rather than teacher implementation. For this purpose, we estimated two models with ASIM scores as the dependent variable and used an F-test to determine the statistical significance of the relationship between ASIM scores and its potential determinants. One model used school-level averages to assess selection at the school level (i.e. whether schools with some characteristics have better implementing teachers). Another model used student level data to assess selection at the student level within schools (i.e. whether students are assigned to classes taught by better implementing teachers based on their characteristics). For both models, F-test *p* values were above .05, which gives confidence that the results of the main analysis are not invalidated by selection bias.

TABLE 7. STUDENT RESULTS OVERALL AND BY STUDENT GROUP, ACT SCIENCE TEST

Category	Effect per unit	Standard error	*p* value	Effect of average ASIM implementation	Effect of average ASIM implementation (percentiles)	Effect of full ASIM implementation
All	1.35	0..13	<.01	0.03	1	0.27
Asian	2.04	0.76	.01	0.05	2	0.40
Black	0.66	0.23	<.01	0.02	1	0.13
Hispanic	1.16	0.50	.02	0.03	1	0.23
Other or unknown	1.58	0.56	<.01	0.04	2	0.31
White	1.64	0.16	<.01	0.04	2	0.32
Female	1.21	0.17	<.01	0.03	1	0.24
Male	1.50	0.18	<.01	0.04	2	0.30
English language learner	1.85	0.97	.06	0.05	2	0.37
Non-English language learner	1.34	0.13	<.01	0.03	1	0.26
Special Education	0.76	0.42	.07	0.02	1	0.15
Non-Special Education	1.39	0.14	<.01	0.04	1	0.27
Free and reduced-price lunch	0.96	0.20	<.01	0.02	1	0.19
Non-free and reduced-price lunch	1.58	0.16	<.01	0.04	2	0.31
Rural	1.04	0.18	<.01	0.03	1	0.20
Urban/suburban	1.76	0.20	<.01	0.05	2	0.35