

RESEARCH REPORT

The Impact of STEM Integration on Student Achievement Using HLM: A Case Study

Danielle Chine^a , Karen Larwin^{b1} 

^aCarnegie Mellon University, USA; ^bYoungstown State University, USA

Abstract:

The study's purpose was to determine the effectiveness of a science, technology, engineering, and math (STEM) program in an urban, Ohio middle school. The definition of integrated STEM education along with curriculum, standards, and implementation models are described. Participants in this longitudinal, quantitative study were all students enrolled in grades 7-8 beginning in 2012-2013 and continuing for seven years. Per grade level for any given year, there were 25-28 participants receiving the treatment by voluntary enrollment in an integrated STEM education program (containing STEM curriculum and applying pedagogies of project-based learning) and 350-425 control participants receiving general education (defined as traditional and lacking both project-based learning and STEM curriculum). We sought to determine if participation in an integrated STEM education program had an impact on student achievement (measured by Ohio State Test scores) and, if any, interaction effects due to gender, socioeconomic status, student race, and attendance rate were present. The use of hierarchical linear modeling (HLM) determined integrated STEM education had a significant, positive effect on achievement combining math and science (participants scoring 31.8 points higher on average) and in science only (participants scoring 38.2 points higher on average) compared to control participants, respectively. No interaction effects were found. These findings pose strong implications for educational leaders in making teacher training and curriculum decisions with the aim of increasing student achievement.

Keywords: STEM education; student achievement; middle school; hierarchical linear modeling; HLM

¹ Corresponding Author: 1 University Plaza, Youngstown, OH, 44555, Email: khlarwin@ysu.edu

To cite this article: Chine, D. R. & Larwin, K. H. (2022). Longitudinal study using hierarchical linear modeling of student and school effects on academic achievement in a middle school STEM program. *Journal of Research in STEM Education*, 8(1), 1-23. <https://doi.org/10.51355/jstem.2022.108>

For the past two decades, the realization of the importance of STEM education in the United States is increasingly gaining momentum from kindergarten into high school and beyond (Bryan & Guzey, 2020; Gonzalez & Kuenzi, 2012; Hansen & Gonzalez, 2014; White, 2014). Fueled by the increase in funding for STEM education initiatives and the decline of U.S. students pursuing STEM degrees, many STEM curriculum companies have been developed with the promise of increasing student achievement (Hess et al., 2015). This work aims to provide a clearer view of the impact and effectiveness of integrated STEM education by expanding on previous similar research (Gabriel et al., 2016, Han et al., 2015; Hansen & Gonzalez, 2014; Hess et al., 2015; Lawanto et al., 2012; Wade-Shepard, 2016).

Defining Integrated STEM Education & Project-based Learning

Currently, there is no single, common definition of integrated STEM education or STEM integration among researchers making it difficult to make progress (Moore et al., 2020). Due to the lack of clarity in the definition of STEM education, the term *integrated STEM education* was designed to encompass the discipline as a whole incorporating all subjects of science, technology, engineering, and math (Giasi, 2018). The emergent discipline has also been defined as interdisciplinary, cross-disciplinary, connected, fused, or transdisciplinary with no definitive boundaries separating each discipline (Honey et al., 2014). The elusiveness of STEM integration has spawned some to define the field as a meta-discipline or the creation of a discipline based on the integration of another discipline into a new whole (Kaufman et al., 2003). Recently, the *Handbook of STEM Education* was published reviewing 109 sources providing definitions and conceptual frameworks of integrated STEM education (Moore et al., 2020). The consensus among researchers on the definition of STEM integration was narrowed down to six common themes, listed below:

STEM integration:

- should be centered around real-world problems,
- applies concepts, principles, and ideas across disciplines,
- frequently uses student-centered learning approaches and peer collaboration,
- requires at least two disciplines,
- can exist on a wide continuum from little (or no) to full integration,
- often contains active learning, student-centered, problem- and project-based teaching pedagogies (Moore et al., 2020).

A wide continuum of STEM integration exists from very little to no integration to full integration containing authentic, project-based learning (Moore et al., 2020). Levels of integration have been defined on a numerical scale (Levels 0-4) with Level 0 containing no integration, Level 3 containing engaging projects, and Level 4 displaying full integration (Burrows & Slater, 2015). More commonly, STEM integration levels have been described nominally as *disciplinary*,

multidisciplinary, interdisciplinary, and transdisciplinary with the latter two levels containing authentic, real-world problems using STEM integration teaching pedagogies (Wang & Knobloch, 2018).

The interdisciplinary and transdisciplinary approach to STEM integration naturally enforces the use of project-based learning (PBL) in the classroom. PBL is a teaching method where students participate in real-world projects often incorporating different subjects. PBL is one of the main approaches to providing STEM education and solidifies the overlap of science, technology, engineering, and mathematics (Hansen & Gonzalez, 2014). PBL utilizes many of the best practice STEM attributes discussed by Moore et al. (2020), such as creating an active and student-centered classroom and serving students with a variety of learning styles. The use of project-based instruction allows students to represent, model, and apply their content knowledge in interesting and unique ways (Wilhelm, 2014). Many researchers agree that PBL learning practices are highly effective as a framework for teaching STEM education (Hansen & Gonzalez, 2014; Moore et al., 2020).

Atkinson (2012) discussed a dichotomy that exists regarding STEM education implementation. Atkinson (2012) reported most stakeholders believe all students should be exposed to STEM education in varying degrees and coined the “Some STEM for All” or “STEM for All” approaches. Gonzalez and Kuenzi (2012) reported that analysts argue that this is not successful. The opposite method argued by Atkinson (2012) to be more effective, particularly at the high school level, is providing intensive STEM education to a limited number of students, based primarily on student interest, and is called the “All STEM for Some” approach. The dichotomy of STEM integration delivery into schools is a controversial topic with more research needed to determine what method is most impactful on students’ success and achievement (Atkinson, 2012). Elrod et al. (2012) argued that Atkinson’s (2012) ideas to concentrate on providing STEM education to fewer students showing interest in the discipline reinforces the exclusivity and disparity of minorities and underrepresented groups. They concluded that Atkinson’s (2012) model of “All STEM for Some” strengthens the STEM pipeline for a select few. Many researchers consider this a major criticism of the “All STEM for Some” approach to providing STEM education (Elrod et al., 2012). This study will add to the current research on the impact of the “All STEM for Some” approach on student achievement.

History of Funding STEM Education

Many historical events led to an increase in STEM funding and the creation of many STEM curriculum companies and programs. World War II and the Soviet Union’s launch of Sputnik 1 increased U.S. research in advancing technologies propelling STEM education efforts (Gonzalez & Kuenzi, 2012; Hansen & Gonzalez, 2014). In 1983, *A Nation at Risk* was published as an educational reform policy reviving the STEM movement in public education (Mahoney, 2009;

Mahoney, 2010). *A Nation at Risk* was extremely influential in the development of national standards produced by the National Research Council (NRC) and the American Association for the Advancement of Science (AAAS) among others. In 2015, the Every Student Succeeds Act (ESSA) was passed taking away some control from the federal government in public education, evident in NCLB, and giving more decision-making latitude to local school districts (Achieve, 2017; ESSA, 2017). This gave local entities the ability and funding to provide formal STEM programming and the ability to purchase technology. Beginning in 2017, \$1.6 billion was given to districts to improve student academic achievement. Using similar funding determinations as Title I, districts can improve the use of technology, including STEM education and digital literacy, to increase student achievement (ESSA, 2017). The increased funding for 21st-century programming has created a demand for high-quality STEM curricula (Achieve, 2017).

Standards & Curriculum

Currently, there are no nationally recognized standards for STEM education. Han and Buchmann (2016) stated the lack of curriculum standardization in the U.S. has contributed to lower science achievement and other STEM disciplines. The lack of continuity and standards has stimulated the NRC and the National Academies of Engineering to highlight the need for intentional and explicit instruction of the STEM curriculum and standards (Radloff & Guzey, 2016). There are *Standards for Technological Literacy (STL)* developed by the International Technology and Engineering Educators Association (ITEEA) (ITEA, 2000/2002/2007). In 2019, the ITEEA developed a plan to revise the standards which have yet to be released. Presently, the most popular STEM and engineering curriculum companies in the United States are Project Lead the Way (PLTW), Engineering by Design (EbD), EverFi, and STEM Education Works. Two of these are reviewed in literature: PLTW and The Infinity Project (Stohlmann et al., 2012) with the former explained in greater detail below and the latter reporting increasing interest in STEM disciplines among students (Orsak et al., 2021).

Project Lead the Way (PLTW)

Due to the increase in national funding allocated to public schools, particularly under ESSA (2015), there has been an influx of middle school pre-engineering and STEM curricula. PLTW focuses on STEM curriculum comprising areas of computer science, engineering, and biomedical science to improve problem-solving, critical thinking, communication, and creativity. PLTW is aligned to the Next Generation Science Standards and the Common Core State Standards for ELA and Math (PLTW, 2021).

The middle school program, PLTW Gateway, consists of 10 units designed to be taught daily for nine weeks (PLTW, 2020). The 10 PLTW Gateway units are, as follows: Design & Modeling, Automation & Robotics, App Creators, Computer Science for Innovators and Makers,

Energy & the Environment, Science of Technology, Magic of Electrons, Medical Detectives, Flight & Space, and Green Architecture. PLTW Gateway curriculum is intended to be taught along with the many components discussed earlier by Honey et al. (2014) as interdisciplinary, cross-disciplinary, and connected with no thorough boundaries separating each discipline. PLTW claims to be student-centered, supportive of various modalities of learning, accommodating to many different learning styles, and able to serve students with disabilities.

Benefits of STEM Integration

The impact of STEM education policies and initiatives on student achievement report varying degrees of success (Dugger, 2010; Gonzalez & Kuenzi, 2012; Snyder, 2018; White, 2014). Gonzalez and Kuenzi (2012) attested there is no single statistic that can fully quantify or encompass the condition of STEM education in the nation and for a variety of reasons the question “what is the condition of STEM education?” may be unanswerable (p. 9). Many researchers note that it is difficult to determine the extent of the impact. Although, from a broad perspective, STEM education has maintained or improved over the past 40 years. Gonzalez and Kuenzi (2012) commented that it is difficult to measure the success of the United States educational system and STEM-related programs due to their complexity. In 2019, the most comprehensive report of the current state of student achievement concerning STEM education was published by the National Science Board (2019). It determined the U.S. ranks in the middle among 19 advanced nations in student performance in the STEM fields with students underperforming compared to students in other developed countries such as Singapore, Taiwan, and South Korea (National Science Board, 2019). However, nationally, U.S. students’ achievement in mathematics has increased in the last 30 years with the largest growth occurring in the first two decades (National Science Board, 2019).

Integrated STEM education has shown many benefits, such as increased student engagement, student learning, and identity development (Bryan & Guzey, 2020; Guzey et al., 2017). Studies report a strong, positive impact particularly in science (Gardner & Tillotson, 2019) and math (English & King, 2019) achievement. The consensus among researchers is that integrated STEM education is successful when all subjects (i.e., science, technology, engineering, math) are interconnected in authentic and meaningful ways (Bryan & Guzey, 2020).

Related Work

PLTW, the nation’s largest non-profit STEM curriculum for middle and high schools (PLTW, 2021) is increasingly used today in classrooms but little research has been conducted on its impact on student achievement and future career choices (Hess et al., 2015). Lawanto et al. (2012) researched the relationship between interest and success expectancy in STEM careers for students taking a PLTW Engineering Design course. Lawanto et al. (2012) determined there was a significant relationship between student self-interest in engineering design and their expectancy

for success with intrinsic motivation being a predictor of future success and STEM career choice. Gabriel et al. (2016) sought to determine the effectiveness of PLTW on the 2016 Missouri Assessment Plan Science scores of 5th- and 8th-grade students. This study included mid-range socioeconomic students only across two school districts comparing test scores of students exposed to the PLTW curriculum and those that were not exposed to the curriculum. They reported a significant, positive effect on student achievement (Gabriel et al., 2016). Last, a systematic review of 31 articles researching the effectiveness of PLTW curricula determined the pros and cons of the program (Hess et al., 2015). PLTW made a positive impact on motivating students into STEM careers, training teachers, and supporting students' interests in STEM. Oppositely, there was little evidence supporting improvement in students' math and science achievement given the high cost of implementation (Hess et al., 2015).

There are three similar research studies found in recent literature to this work. First, Wade-Shepard (2016) investigated the effect of middle school STEM curriculum on both science and math achievement scores. The research was conducted among four Tennessee schools with seventh and eighth grade students participating in the Tennessee Comprehensive Assessment Program. Wade-Shepherd (2016) found a significant, strong, and positive correlation between math and science test scores of students participating in STEM classes compared to those that were not taking STEM classes across the four schools (Wade-Shepard, 2016). The work did not include data analysis of other moderators of achievement, such as attendance, demographic information, and teacher efficacy.

Hansen and Gonzalez (2014) investigated the relationships between STEM learning principles, such as PBL, and student achievement in math and science. This mixed-methods study included middle school students in North Carolina and used a combination of quantitative state assessment and qualitative student survey data. They determined specific STEM practices were associated with performance gains in math and science. For example, projects and science experiments were associated with higher scores in science, and the use of technology and computers were associated with higher scores in math. In addition, these significant and positive correlations were also found among racial minorities (Hansen & Gonzalez, 2014).

The third investigation analyzed both STEM curriculum and PBL strategies on student mathematics performance disaggregated by the low, middle, and high achieving students to determine the degree of effect as a function of student achievement level (Han et al., 2015). The study took place in Texas among three high schools with students in the treatment group participating in STEM PBL activities once every six weeks over the course of three years. This work was similar to our investigation as it used hierarchical linear modeling (HLM) to determine the effect of STEM PBL activities on students' mathematics scores accounting for student moderators such as student SES and race. Han et al. (2015) concluded lower achieving students showed a statistically significant higher rate of growth on math scores compared to middle and

high performing students over the course of three years. They also found student race and SES were strong predictors of student academic achievement (Han et al., 2105). Our proposed work is an amalgam of these prior studies that hope to enrich the research on effective STEM integration.

Significance of the Study

The results of this investigation are highly influential and beneficial to educational administrators and policymakers to gain insight into the academic gains of providing integrated STEM programs. This research will provide educational leaders with a clearer view of the impact and effectiveness of STEM integration on student achievement expanding on previously discussed research (Gabriel et al., 2016, Han et al., 2015; Hansen & Gonzalez, 2014; Hess et al., 2015; Lawanto et al., 2012; Wade-Shepard, 2016). This will assist them with making future educational and fiscal decisions regarding the planning, purchasing, and implementing of STEM programs, particularly at the middle school level. In addition, this research will benefit students as it will help determine the benefits of STEM education programs on future success on state assessments and on providing mastery experiences.

Research Questions

The current investigation seeks to answer the following research questions:

1. Do students participating in an integrated middle school STEM program demonstrate differences in academic achievement in math and science compared to students participating in a traditional general education setting?
2. Do students participating in the integrated middle school STEM program demonstrate differences in achievement due to the interaction effect of gender, race, socioeconomic status (SES), or attendance rate?

Methods

Participants and Setting

The setting for the research investigation was a mid-sized urban district located in northeastern Ohio. During the research investigation, the district was largest in student enrollment compared to any other school district in the county with 4,437 students enrolled in grades kindergarten through 12th grade for the 2018-2019 academic year (ODE, 2019). Both the percent of students in the district economically disadvantaged and the minority enrollment have remained steady for the study duration. The following demographic information was not reported due to numbers being too small: American Indian/Alaskan Native, Asian/Pacific

Islander, English Learner, and Migrant. Table 1 compares middle school demographics from the first to last year of the study.

Table 1.
Middle School Demographics from First to Last Year

Demographics	2012-2013	2018-2019
Black, Non-Hispanic	10.3%	12.8%
Hispanic	3.8%	6.2%
Multiracial	2.2%	6%
White, Non-Hispanic	82.9%	73.9%
Students with Disabilities	12.8%	18%
Economic Disadvantage	47.0%	52.5%
Total students	1,315	1,061

The participants were all students in grades 6-8 enrolled from 2012-2013 to 2018-2019. There were 25-28 students participating in the integrated STEM program each year for 7th and 8th grade. The number of general education students ranged from 350-425 students per grade level depending on the school year. Control participants received general education defined as traditional and lacking both STEM curriculum and project-based learning. Participants had the subjects of math, science, ELA, and social studies each taught by a different teacher and in isolation. Participants had nine week or semester rotations (depending on year and grade level) of elective courses (i.e., physical education, art, technology, general music). Band and choir were available all year and took the place of an elective course.

Participants in the treatment group were those who volunteered to participate in an integrated STEM program. Participants received all 10 PLTW Gateway modules (7th and 8th grade combined) taught as nine-week units except for App Creators and Science of Technology which were instructed in parts throughout the year to align with standards taught in other subjects. There were 25-28 participants for 7th or 8th grade with one teacher teaching science and math and another teaching science and social studies, while both taught the select PLTW modules. In addition, participants completed collaborative, real-world projects, such as the Soap Box Derby requiring the building and testing of Soap Box Derby cars and a weather balloon launch applying concepts learned in the Flight and Space PLTW module using the teaching pedagogies of active and project-based learning.

Reliability and Validity of the Instrument

Student achievement was measured using Ohio State Test (OST) data as instruments and reported by the Ohio Department of Education (ODE). Student demographic data were collected from the Education Management Information System (EMIS) database. Student demographic information consisted of student race, defined as American Indian or Alaskan Native, Asian or Pacific Islander, Multiracial, Hispanic, Black (non-Hispanic), and White (non-Hispanic). Both race and SES were self-reported by parents and guardians to the school district. Student race was self-reported by parents and guardians with no required documentation. SES was derived from free-and reduced-lunch status reported by family group W-2 forms and was considered a valid metric for assessment of SES.

The student measure of academic achievement was determined using annual OST scores taken each spring by students in grades three through eight for all students in the state of Ohio. Testing is mandatory in grades three and above with particular tests by subject required at the high school level with an appropriate score required for graduation. Students took the mathematics OST both their 7th and 8th-grade year and science in 8th grade.

The reliability and validity of Ohio state assessment data are reported by the ODE annually and are considered one of the stronger instruments used for defining achievement. State assessments are norm-referenced and standardized to ensure alignment with Ohio's Learning Standards for each grade level and subject. Reliability of all Ohio Academic Achievement (OAA) assessments (administered in 2012-2013 and 2013-2014) ranged from 0.87-0.90 using Cronbach's alpha and the standard error of measurement (SEM) ranged from 10.24-13.03, respectively (ODE, 2014). Reliability of the American Institute of Research (AIR) assessments across subjects (administered from 2015-2016 to 2018- 2019) ranged from 0.90-0.94 using Cronbach's alpha and the SEM ranged from 9.81-15.49 (ODE, 2020). Reliability measures for the 2014-2015 administration of the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment were not released due to only being administered for one year. The reported Cronbach's alpha measures for calculating internal consistency are all well above 0.70 indicating all administered assessments have strong reliability. There was the suspension of state testing for the 2019-2020 school year due to the coronavirus crisis which caused school closure and suspended state testing. Therefore, no student test data were collected for the 2019-2020 academic school year.

Procedures & Analysis

As a causal-comparative investigation, data from the years 2010-2011 through 2018-2019 were collected. This was due to 7th-grade students in the first year (2012-2013) having 5th-grade science scores used in the analysis in the comparison of 5th to 8th-grade science achievement.

Once the Youngstown State University IRB approved the use of the data, the county educational offices extracted the data for the researchers.

The method of data analysis used was multilevel modeling. Multilevel modeling is used in many fields of study particularly in education, social work, health, business sectors, and the social sciences (Woltman et al., 2012). This type of modeling is known by several names, such as hierarchical linear-, mixed level-, mixed-effects-, random effects-, random coefficient (regressions), and (complex) covariance components- modeling (Raudenbush & Bryk, 2002). Multilevel modeling and HLM are complex forms of ordinary least squares (OLS) regression and are used when predictor variables are at different hierarchical levels to determine the variance within the outcome variables. HLM is primarily used for creating statistical models of variables that depend on more than one level, or nested data. HLM simultaneously determines relationships within and among hierarchical levels within data sets thereby making it an effective method of calculating variance among variables at varying levels than other statistical analysis techniques. HLM is becoming an increasingly popular method of advanced statistical analysis due to advancements in statistical theory and statistical modeling programs (Woltman et al., 2012).

Table 2 displays the proposed factors at each hierarchical level that affect students' achievement with variables' names used in the HLM and SPSS software programs.

Table 2.

Proposed Factors at Each Hierarchical Level that Affect Students' Achievement

Hierarchical Level	Category	Variables	HLM Variable Code
Level-2	School Level	Participation in a STEM program	STEMMARK
		Participation in the general education setting	
		Assessment Type (Grade and school year)	ASSMTTYP
Level-1	Student Level	5 th - and 8 th - grades' science tests	GRADE
		OST scaled score	SCALEDSC
		Gender	GENDER
		Race	RACE
		SES	SES
		Attendance	ATTEND

Results

HLM was used to analyze OST data for students in grades five through eight to determine the effect of student achievement, the outcome variable, as a function of varying hierarchical levels. The student-level variables are listed with the labels in parenthesis and corresponding variable coding given, as follows:

- Socioeconomic status (SES); economically disadvantaged=1, not economically disadvantaged=0
- Gender (GENDER); female=1, male=0
- Race (RACE); White (Non-Hispanic)=7, Puerto Rican= 6, Multiracial=5, Hispanic=4, Black (Non-Hispanic)=3, Asian=2, Alaskan Native/ Am. Indian=1
- Attendance (ATTEND); Between or equal to 0 and 1. Coded as percent attendance at a decimal rate.

The two school-level variables of grade level and STEM program designation were given the *SPSS* labels and variable coding. Grade (GRADE) was given grade 5=5, grade 6=6, and grade=7. Student participation in integrated STEM programming (STEM) was indicated by 1 and student participation in a general education setting was indicated by 0.

As indicated above, OST results are the outcome variable for this investigation. The student level OST data was collected for each academic year beginning in the school year 2010-2011 and concluding in 2018-2019. Table 3 displays student gender disaggregated by STEM participation for the period included in this investigation.

Table 3.
Student Gender by STEM Participation for the years 2010-2011 through 2018-2019

	Gender	Frequency	Percent
Gen. Ed. (<i>n</i> = 3035)	Male	1515	49.9
	Female	1520	50.1
	Total	3035	100
STEM (<i>n</i> = 205)	Male	123	60
	Female	82	40
	Total	205	100

The participants included 3,035 general education students and 205 STEM students. The general education students were 50.1 percent female and STEM students only 40 percent female with a ratio of male to female of 3:2.

Among the 3,032 general education participants spanning the study years, the majority of students were White (77.3%), with the second race indicated as Black (13.1%). Other races were represented at significantly lower percentages. Among the 205 STEM participants who provided information on race, the majority were also White (87.3%) with other races, such as Multiracial (6.3%) and Black (3.9%), representing a much lower percentage. The STEM program was less racially diverse than the general education population.

HLM Model 1- Academic Achievement by Year

The use of HLM to determine the effect of integrated STEM programming on student achievement was modeled using different variables at level-1 and level-2. The first model used OST score (SCALEDSC) and STEM participation (STEMMARK) at level-1 and the assessment type indicated by year (ASSMTTYP) as level-2. Table 4 displays level -1 descriptive statistics.

Table 4.

Model 1: Level-1 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
SCALEDSC	8874	598.94	149.06	314.00	868.00

There were 8,874 data points at level-1 with a mean scaled score of 598.94. The test scores are mutually exclusive for individual subjects, grades, and years but HLM accounts for this by nesting the data within level-2. There were nine assessment types indicated by years included in level-2. Equation 6 displays the HLM equation at level-1, OST scores are shown as the outcome variable ($SCALEDSC_{ij}$) and STEM participation ($STEMMARK_{ij}$) is the level-1 predictor variable. Equation 7 shows level-2 with assessment type ($ASSMTTYP_j$) as the level-2 predictor variable.

Level-1 Model

$$SCALEDSC_{ij} = \beta_{0j} + \beta_{1j}*(STEMMARK_{ij}) + r_{ij} \quad (6)$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(ASSMTTYP_j) + u_{0j} \quad (7)$$

$$\beta_{1j} = \gamma_{10}$$

The mixed model below (Equation 8) substitutes the intercept of the j th level-2 (β_{0j}) from Equation 7 into Equation 6 to get the mixed model shown below (Equation 8). The combined model contains both the level-1 and level-2 predictors and a term across levels containing both random and fixed effects unique to HLM analysis. The Microsoft Office User analysis of variance (ANOVA) model was used to determine the mean achievement scores among both general education students and students participating in a STEM program and compare the differences. This was performed to measure the variation between student-level and grade-level assessment groups. This mixed model, combining both fixed and random effects, was used to analyze the relationship between student achievement as a function of STEM programming versus general education programming.

Mixed Model

$$\text{SCALEDSC}_{ij} = \gamma_{00} + \gamma_{01} * \text{ASSMTTYP}_j + \gamma_{10} * \text{STEMMARK}_{ij} + u_{0j} + r_{ij} \quad (8)$$

The final estimation of fixed effects with robust standard errors for Model 1 is shown in Table 5.

Table 5.

Model 1: Final Estimation of Fixed Effects (with Robust Standard Errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	836.77	99.32	8.43	7	<0.001
ASSMTTYP, γ_{01}	-211.11	49.72	-4.25	7	0.004
For STEMMARK slope, β_1					
INTRCPT2, γ_{10}	31.33	2.29	13.69	8864	<0.001

Fixed effects were used because the level-2 group was a unique entity and j was small indicating the number of years ($j < 10$). Robust standard errors were used for both Model 1 and Model 2 due to confidence in the distribution of the dependent variable of assessment type at level-2. The overall mean intercept adjusted for student achievement by year (ASSMTTYP) for STEM students (STEMMARK) was determined to be 31.3 (INTRCPT2, γ_{10}), indicating a significant correlation between STEM participation and student achievement. Student achievement as indicated by OST scores for a given year, grade, and subject indicate that a STEM student was predicted to score 31.3 points higher than general education students. All p-values

are statistically significant ($p < 0.004$) supporting the correlation between STEM program participation and student achievement.

The final estimation of variance shown in Table 6 displays the random error associated with the use of the final estimation of fixed effects. The random effect at level-1 has a standard deviation of 44.68.

Table 6.

Model 1: Final Estimation of Variance Components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	131.48	17285.64	7	74797.10	<0.001
level-1, r	44.68	1996.36			

HLM Model 2- Comparing 5th- and 8th-Grade Science

A second model examines whether participating in an integrated middle school STEM program demonstrates differences in academic achievement in science?

The goal of the second model was to effectively predict the OST score for students taking both the fifth grade and eighth grade science OST tests as a function of STEM participation. The model used OST score (SCALEDSC) and STEM participation (STEMMARK) at level-1 (similar to Model 1) and the assessment type indicated by year (GRADE) as level-2. Table 7 displays descriptive statistics for level-1.

Table 7.

Model 2: Level-1 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
SCALEDSC	4048	562.50	156.94	314.00	868.00

There were 4,048 data points at level-1 with a mean scaled score of 562.50. The test scores are mutually exclusive for individual grades and years but HLM accounts for this by nesting the data within level-2. The descriptive statistics at level-2 are shown in Table 8.

Table 8.

Model 2: Level-2 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
GRADE	9	6.00	1.50	5.00	8.00

There were nine science assessments indicated by the variable GRADE occurring over the course of the study beginning in 2010-2011 and ending in 2018-2019. The minimum value was 5 indicating fifth grade and the maximum value was 8 indicating eighth grade.

Equation 9 displays the HLM equation at level-1, OST scores are shown as the outcome variable (SCALEDSC_{ij}) and STEM participation ((STEMMARK_{ij}) is the level-1 predictor variable. Equation 10 shows level-2 with science assessment (GRADE_j) as the level-2 predictor variable.

Level-1 Model

$$\text{SCALEDSC}_{ij} = \beta_{0j} + \beta_{1j} * (\text{STEMMARK}_{ij}) + r_{ij} \quad (9)$$

Level-2 Model

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} * (\text{GRADE}_j) + u_{0j} \\ \beta_{1j} &= \gamma_{10} \end{aligned} \quad (10)$$

The mixed model in Equation 11 substitutes the intercept of the j th level-2 (β_{0j}) from Equation 10 into Equation 9 to get the mixed model shown below (Equation 11). This model was found to be significant in predicting student achievement as a function of the defined level-1 and level-2 variables.

Mixed Model

$$\text{SCALEDSC}_{ij} = \gamma_{00} + \gamma_{01} * \text{GRADE}_j + \gamma_{10} * \text{STEMMARK}_{ij} + u_{0j} + r_{ij} \quad (11)$$

The final estimation of fixed effects with robust standard errors for Model 2 is shown in Table 9.

Table 9.

Model 2: Final Estimation of Fixed Effects (with Robust Standard Errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	170.67	151.51	1.13	7	0.297
GRADE, γ_{01}	68.17	18.98	3.59	7	0.009
For STEMMARK slope, β_1					
INTRCPT2, γ_{10}	38.19	3.14	12.17	4038	<0.001

The overall mean intercept adjusted for student achievement by year (ASSMTTYP) for STEM students (STEMMARK) was determined to be 38.2 (INTRCPT2, γ_{10}), indicating a

significant correlation between STEM participation and student achievement as evidenced by fifth grade to eighth grade science OST scores. Student achievement measured by OST scores for fifth and eighth grade science predicted STEM students will score 38.2 points higher than general education students. This correlation was stronger than all assessment types used in Model 1. All p-values are statistically significant ($p < 0.009$) supporting the correlation between STEM program participation and student achievement in science except for INTRECPT2, γ_{00} with a p-value of 0.297.

The final estimation of variance shown in Table 10 displays the random error associated with the use of the final estimation of fixed effects. The standard deviations in the random effect associated with level-1, r was 45.42 and an INTRCPT1, u_0 of 128.88.

Table 10.

Model 2: Final Estimation of Variance Components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, u_0	128.88	16608.96	7	25720.24	<0.001
level-1, r	45.42	2062.89			

Additional analyses sought to determine if students participating in an integrated middle school STEM program demonstrated differences in academic achievement due to student gender, race, SES, and attendance. Out of the 3,237 and 3,240 students with data for GENDER and RACE, the Pearson bivariate correlation was extremely low ($r = 0.069$ and -0.049 , respectively) indicating a lack of practical significance. Therefore, there is no practical significance between these student-level variables of race and gender and STEM participation.

Preliminary bivariate relationships between STEM participation and student-level variables of gender, SES, race, and attendance were determined using Pearson correlation coefficients. There was a very small correlation between attendance rates with Pearson r values of 0.102. All other Pearson r values were extremely low (ranging from -0.080 to 0.069) indicating no practical significance in the relationship or correlation between these student-level variables and STEM participation.

Discussion

Impact on Math & Science Achievement

Model 1 analyzed student achievement on both math and science assessments combined using OST score and STEM participation at level-1 and the assessment type by year at level-2.

The singular level-2 variable clustered the level-1 participants into nine groups ($n=9$) for each tested year beginning in the school year 2010-2011 through 2018-2019 creating a longitudinal sample analysis. A mixed model, containing both fixed and random effects, combined both the level-1 and level-2 predictors and was found to be significant in predicting student achievement as a function of STEM participation for a given tested year cluster. The overall mean intercept adjusted for student achievement by year for STEM students was determined to be 31.3 points, indicating a significant correlation between STEM participation and student achievement in both math and science combined. Student achievement as indicated by OST scores for a given year, grade, and subject indicate that a STEM student is predicted to score 31.3 points higher than general education students. All p-values were significantly small ($p<0.004$) supporting the correlation between STEM program participation and both math and science achievement.

Model 1 predicts students participating in the integrated STEM program will score significantly higher on achievement assessments for both math and science combined compared to general education peers scoring 31.3 points higher on average. This finding aligns with Wade-Shepard (2016) which found a significant, strong, and positive correlation between math and science achievement and students participating in integrated STEM education coursework. However, the systematic review discussed previously by Hess et al. (2015) found little evidence supporting improvement in students' math and science achievement given the high cost of implementation.

Impact on Science Achievement

Students took the OST science tests in 5th and 8th grade only. Because the STEM program began in the 7th grade, students in the treatment group had two years of STEM programming by the time they took the 8th-grade test. The use of HLM to determine student achievement as a function of STEM integration was powerful when comparing scores from the 5th to 8th grades.

Model 2 used OST score and STEM participation at level-1 (similar to Model-1) and the science assessment type indicated by year at level-2. The singular level-2 variable clustered the level-1 participants into nine groups ($n=9$) for each tested year of science only beginning in the school year 2010-2011 through 2018-2019 creating a longitudinal sample analysis. A mixed model, containing both fixed and random effects, combined both the level-1 and level-2 predictors and was found to be significant in predicting student achievement as a function of STEM participation for a given tested year cluster. The overall mean intercept adjusted for student achievement by year for STEM students was determined to be 38.2, indicating a significant correlation between STEM participation and student achievement in science. Student science achievement as indicated by OST scores for a given year, grade, and subject indicate that a STEM student is predicted to score 38.2 points higher than general education students. All p-values are statistically

significant ($p < 0.009$) supporting the correlation between STEM program participation and science achievement except for INTRECPT2, γ_{00} with a p-value of 0.297.

The predictive results of Model 2 indicate through comparison of descriptive statistics and HLM analysis, that middle school students participating in integrated STEM programming scored significantly higher on the OST in science compared to their general education peers scoring above 38.2 points higher on average. The impact of STEM participation on student achievement was stronger when comparing science only compared to both math and science achievement combined. In particular, the positive impact on science is supported in the literature (Gardner & Tillotson, 2019). In addition, our finding may support the concept proposed by Hansen and Gonzalez (2014) that specific STEM practices may be associated with performance gains in math and science. For example, projects and science experiments may be associated with higher scores in science compared to math (Hansen & Gonzalez, 2014). This hypothesis would support our findings as the integrated STEM program used PBL as a program foundation. In addition, Hansen and Gonzalez (2014) found significant and positive correlations among racial minorities in science achievement. This study did not find any differences in achievement due to interaction effects, such as demographics (i.e., student race). More research needs to be conducted to determine the effects of integrated STEM education on science achievement (and math) among racial minorities.

Limitations

There were a few limitations related to both the data collection process and research methodology. Several students were missing SES status and were falsely identified as not economically disadvantaged. This significantly lowered the percent range of economically disadvantaged students from 38.8% to 47.3% with the state reporting 51% to 54% (ODE, 2019). Another limitation in data collection was the absence of attendance data for several students and the self-reporting of student race data by parents. Student mortality was insignificant with only one to two students choosing to leave the treatment program by grade level each year. Therefore, the migration of students out of the program posed negligible internal validity and reliability threats. Limitations related to the methodology were generalizability due to limited external validity, as in most case study research, however, longitudinal duration of seven years provided a clearer view of the impact of STEM programming on the specific population. There were a limited number of teachers providing STEM programming making it hard to discern if the differences in achievement between STEM and general education students were due to STEM programming or teacher-related factors, such as self-efficacy. Fortunately, OST scores were used as a standardized measure to assess student learning strengthening internal validity. The primary limitation is selection bias due to voluntary participation in the STEM program with many STEM students demonstrating higher than average achievement prior to participating in the program.

This is evidenced by the mean OST scores in science for fifth grade STEM students being larger than the mean of general education students for all tests. HLM analysis accounts for much of this bias as it analyzes nested data by calculating group variances at each hierarchical level. However, there is a possible validity threat due to students participating in the STEM program treatment possibly displaying greater motivation and interest to learn. Therefore, the impact of STEM programming may be more difficult to determine due to a degree of validity threat from selection bias.

Implications for Practice

The findings indicate integrated STEM programming has a positive impact on middle school students and increases both science and math performance. This has powerful implications for educational leaders, particularly with changes in legislation that gave more power to local school districts and increased funding under ESSA (Achieve, 2015; ESSA, 2015). Therefore, it is important educational leaders are aware of the impact integrated STEM programming and PBL has on student achievement. In addition, due to the increase in funding for STEM education, there are many STEM curriculum companies promising quality technological curriculum, such as PLTW. This research will assist educational leaders in making the appropriate financial decisions regarding the purchasing of effective STEM programming and professional development for teachers.

There were two implementation dichotomies discussed previously: isolation versus integration and “Some STEM for All” versus “All STEM for Some”. This research further demonstrates the success of STEM on student achievement using the integration model which is supported by many in the previous research literature (Dugger, 2010; Sanders, 2009). In addition, this research analyzed the success of a STEM program model utilizing the “All STEM for Some” approach described and supported by Atkinson (2012) and Elrod et al. (2012), particularly at higher grade levels.

The argument among researchers is not whether integrated STEM education is effective at the elementary, middle, or high school level, but rather which grade level is most impactful for the introduction of STEM practices on student achievement, and subsequently, future success. Both the elementary and high school years have been shown to be powerful regarding shaping students’ perceptions of their learning and impact on the development of integrated STEM education practices. Somewhere between the two may be the “goldilocks” zone, the middle school years, when students are beginning to develop attitudes and beliefs regarding their abilities in STEM and possibilities of future careers (Christensen et al., 2015). This research supports the positive impact STEM programming has during the middle school years.

The National Middle School Association (NMSA, 2010) recommended integrated STEM curriculum and instruction at the middle school level as it offers engaging and holistic instruction

for all learners with studies finding the integration of mathematics and science having a positive influence on students' attitudes toward school, their motivation to learn, and academic performance. Middle school is a pivotal time for students as their viewpoints on education and careers are greatly impacted by their environment and their focus shifts to future careers. This research further supports the importance of middle school STEM in shaping students' perspectives on future careers (Moreno et al., 2016).

Direct implications of this work show the importance and positive impact of integrated STEM initiatives on student academic achievement. However, the study also highlights the lack of STEM interest among racial minorities and females. Students volunteering to participate in the integrated STEM program were less racially diverse than the general population. In addition, the ratio of male to female students participating in the STEM program was 3:2 (general population 1:1). This may support the conclusion of Gonzalez and Kuenzi (2012) stating "All STEM for Some" approaches reinforce exclusivity and increase racial disparity among minorities and underrepresented groups. Stakeholders need to be aware of such disparities especially as years of implementation pass with little to no increase in racial diversity and gender equality among participation in STEM-related programs.

Considerations for Future Practice and Research

There are several considerations regarding directions for future practice and research. An extension of this investigation would be to determine the differences in OST scores among low, middle, and high achieving students as a function of STEM participation. Han et al. (2015) found low achieving students exposed to STEM programming experienced the most growth in achievement compared to their middle and higher achieving peers. Disaggregating student achievement by performance level would gain deeper information into the impact of STEM on different levels of learners. Other considerations would be to determine teachers' self-efficacy for the few educators providing the STEM programming and gain a deeper grasp of the quality of STEM and PLTW training. An expansion upon this would be to determine the technological pedagogical knowledge of the teachers providing STEM programming to determine their individual impact on student achievement. Han et al. (2015) conducted an analysis determining the different types of content knowledge necessary to be a successful STEM educator. This model called the Technological Pedagogical Content Knowledge (TPACK) framework assesses teachers' self-efficacies to determine their impact on student achievement. Collecting TPACK information and survey data would strengthen the results of the impact integrated STEM education has on student achievement.

Conclusion

This longitudinal study reported OST scores along with demographic factors such as gender, SES, race, and attendance to determine the impact of STEM programming on specific populations of middle school students. The use of multilevel, statistical analysis through HLM determined integrated STEM programming had a significant, positive effect on student achievement in both math and science, and an even stronger impact isolating science achievement by itself. The predictive results of HLM analysis determined STEM students scored significantly higher on the OST in science and math combined, scoring 31.8 points higher on average and 38.2 points higher in science compared to their general education peers. No interaction effects were determined between STEM participation and gender, SES, student race, and attendance rate. More research needs to be conducted on the impact of teacher efficacy, individual classroom influences, and quality of STEM training for educators. The results indicate that integrated STEM programming in middle school has a positive effect on student achievement, however, when implementing volunteer-based STEM programs stakeholders must be aware of reinforcing racial disparities and gender inequalities.

References

- Achieve. (2017). Leveraging ESSA to promote science and STEM education in states. https://www.achieve.org/files/Achieve_STEMreport7.12.17.pdf
- Atkinson, R. D. (2012). Why the current education reform strategy won't work. *Issues in Science and Technology*, 28(3). 29-36.
- Bryan, L., & Guzey, S. S. (2020). K-12 STEM Education: An overview of perspectives and considerations. *Hellenic Journal of STEM Education*, 1(1), 5-15.
- Burrows, A., & Slater, T. F. (2015). A proposed integrated STEM framework for contemporary teacher preparation. *Teacher Education & Practice*, 28(2/3), 318-330.
- Christensen, R., Knezek, G., & Tyler-Wood, T. (2015). Alignment of hands-on STEM engagement activities with positive STEM dispositions in secondary math students. *Journal of Science Education and Technology*, 24(6). 898-909.
- Dugger, W. E. (2010). Evolution of STEM in the United States. *International Technology and Engineering Educators Association*.
- Elrod, S., Sawyer, K., & Van Lint, V. (2012). Better STEM for all. *Issues in Science and Technology*, 28(4). 11-15.
- English, L. D., & King, D. (2019). STEM integration in sixth grade: Designing and constructing paper bridges. *International Journal of Science and Mathematics Education*, 17(5), 863-884.
- Gabriel, C., Quinlin, L., & Smith, L. (2016). *The Effect on Standardized Science Test Scores, for Fifth and Eighth Grade Students, in Mid-range Socioeconomic Schools, with Exposure to Project Lead the Way Curriculum* [Doctoral dissertation, Northwest Missouri State University].

- Gardner, M., & Tillotson, J. W. (2019). Interpreting integrated STEM: Sustaining pedagogical innovation within a public middle school context. *International Journal of Science and Mathematics Education*, 17(7), 1283-1300.
- Giasi, T. (2018). *Post-Program Implementation of Integrated STEM Instruction: A Qualitative Case Study* [Doctoral Dissertation, The Ohio State University].
- Gonzalez, H. & Kuenzi, J. (2012). Science, technology, engineering, and mathematics (STEM) education: A primer. *Congressional Research Service*, 7(5700), 1-34. <https://fas.org/sgp/crs/misc/R42642.pdf>
- Guzey, S. S., Harwell, M., Moreno, M., Peralta, Y., & Moore, T. J. (2017). The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics. *Journal of Science Education and Technology*, 26(2), 207-222.
- Han, S. & Buchmann, C. (2015). Aligning science achievement and STEM expectations for college success: A comparative study of curriculum standardization. *The Russell Sage Foundation Journal of the Social Sciences*, 2(1). 192-211. <https://doi.org/10.7758/rsf.2016.2.1.09>
- Hansen, M. & Gonzalez, T. (2014). Investigating the relationship between STEM learning principles and student achievement in math and science. *American Journal of Education*, 120(2). 139-171. <https://doi.org/10.1086/674376>
- Hess, J. L., & Sorge, B., & Feldhaus, C. (2016, June), The Efficacy of Project Lead the Way: A Systematic Literature Review Paper presented at 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana. <https://doi.org/10.18260/p.26151>
- Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Committee on Integrated STEM Education; National Academy of Engineering; National Research Council.
- ITEA (2000/2002/2007). Standards for technological literacy: Content for the study of technology. ITEEA.
- Kaufman, D., Moss, D. M., & Osborn, T. A. (Eds.). (2003). *Beyond the boundaries: A transdisciplinary approach to learning and teaching*. Greenwood Publishing Group.
- Lawanto, O., Santoso H. B., & Liu, Y. (2012). Understanding of the relationship between interest and expectancy for success in engineering design activity grades 9-12. *Journal of Educational Technology & Society*, 15(1). 152-161.
- Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). STEM integration: A synthesis of conceptual frameworks and definitions. In *Handbook of research on STEM education* (pp. 3-16). Routledge.
- Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology*, 25(6). 889-898. <https://doi.org/10.1007/s10956-016-9643-3>
- National Middle School Association. (2010). This we believe: Keys to educating young adolescents. *National Middle School Association*.
- National Science Board (2019). Science and Engineering Indicators 2020. Elementary and Secondary Mathematics and Science Education. NSB-2019-6. Alexandria, VA. <https://nces.nsf.gov/pubs/nsb20196>

- Ohio Department of Education (2014). Ohio Achievement Assessments Statistical Summary. July 2014. <http://education.ohio.gov/Topics/Testing/Testing-Results/Results-for-Ohios-State-Tests>
- Ohio Department of Education (2019). [unknown] School Report Card. Ohio School Report Cards. <https://reportcard.education.ohio.gov/district/overview/048298>
- Ohio Department of Education (2019). "Understanding Ohio's State Test Reports 2018-2019". <http://education.ohio.gov/Topics/Testing/Testing-Results/Results-for-Ohios-State-Tests>
- Ohio Department of Education (2019). Ohio's State Tests in English Language Arts, Mathematics, Science, and Social Studies. <http://education.ohio.gov/Topics/Testing/Ohios-State-Test-in-ELA-Math-Science-SocialStudies>
- Orsak, G., & Willis, B. (2021, November), *The Infinity Project – Engineering for the High School Classroom* Paper presented at 2003 GSW, unknown. 10.18260/1-2-620-38521
- Project Lead the Way. (2021). <https://www.pltw.org/our-programs>
- Radloff, J. & Guzey, S. (2016). Investigating preservice STEM teacher conceptions of STEM education. *Journal of Science Education and Technology*, 5(25). 759-774. <https://doi.org/10.1007/s10956-016-9633-5>
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Sanders, M. (2009). STEM, STEM education, STEMmania. *The Technology Teacher*, 1(1). 20-26.
- Snyder, M. (2018). A century of perspectives that influenced the consideration of technology as a critical component of STEM education in the United States. *Journal of Technology Studies*, 44(2), 42-57.
- Stohlmann, M., Moore, T., & Roehrig, G. (2012). Considerations for teaching integrated STEM education. *Journal of Pre-College Engineering Education Research (J-PEER)*, 2(1), 28-34. <https://doi:10.5703/1288284314653>
- Wade-Shepard, A. A. (2016). The effect of middle school STEM curriculum on science and math achievement scores. [Doctoral dissertation, Union University].
- Wang, H. H., & Knobloch, N. A. (2018). Levels of STEM integration through agriculture, food, and natural resources. *Journal of Agricultural Education*, 59(3), 258-277. <https://doi.org/10.5032/jae.2018.03258>
- White, D. W., (2014). What is STEM education and why is it important? *Florida Association of Teacher Educators Journal*, 1(14), 1-9.
- Wilhelm, J. (2014). Project-based instruction with future STEM educators: An interdisciplinary approach. *Journal of College Science Teaching*, 43(4), 80-90.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69