www.ijres.net

Using HLM to Determine a STEM Programs Impact on Middle School Academic Achievement

Danielle R. Chine
Carnegie Mellon University, United States

Karen H. Larwin
Youngstown State University, United States

To cite this article:

Chine, D. R. & Larwin, K. H. (2022). Using HLM to determine a STEM programs impact on middle school academic achievement. *International Journal of Research in Education and Science (IJRES), 8*(1), 1-17. https://doi.org/10.46328/ijres.2607

International Journal of Research in Education and Science
2022, Vol. 8, No. 1, 1-17 https://doi.org/10.46328/ijres.2607

Using HLM to Determine a STEM Programs Impact on Middle School Academic Achievement

Danielle R. Chine, Karen H. Larwin

Article Info

Article History

Received:
21 August 2021
Accepted:
23 December 2021

Keywords

Hierarchical linear modeling
Multivariate analysis
Student achievement
Multilevel modeling
STEM education

Abstract

Hierarchical linear modeling (HLM) has become an increasingly popular multilevel method of analyzing data among nested datasets, in particular, the effect of specialized academic programming within schools. The purpose of this methodological study is to demonstrate the use of HLM to determine the effectiveness of STEM programming in an Ohio middle school. This longitudinal study analyzes potential moderators of gender, socioeconomic status, student race, and attendance rate along with state test scores to quantify achievement. HLM determined integrated STEM education had a significant, positive effect on achievement in math and science combined (students scoring 31.8 points higher on average) and science achievement (students scoring 38.2 points higher on average) compared to traditional education students, respectively. There were little to no interaction effects determined between STEM participation and student factors. This demonstrates HLM as a powerful statistical tool for quantifying the impact of academic programming on student achievement.

Introduction

The use of hierarchical linear modeling (HLM) has become an increasingly popular method of analyzing data among nested datasets, in particular, in determining the effect of specialized academic programs offered in specific classrooms nested within schools. Multilevel modeling and HLM is used in many fields of study specifically 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).

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. Although, from a broad perspective, STEM education has maintained or improved its impact over the past 40 years. Gonzalez and Kuenzi (2012) comment it is difficult to measure the success of the United States educational system due to its complexity. Despite these challenges, the use of HLM to determine the impact of academic programming, such as STEM education programs is a powerful tool for educators.

Literature Review

The use of HLM is the preferred method of analyzing multilevel datasets as it accounts for the communal variance fundamental to hierarchical datasets. Basic linear regression methods, disaggregation, and aggregation, which did not consider hierarchical data and their shared variances, were used by researchers before the use of HLM. Although both of these outdated methods made analyzing hierarchical data possible they created other problems. These included the wrong variances assigned between variables, data dependencies, and an increased probability of a Type I error (Woltman et al., 2012). Disaggregation of data ignores the differences in hierarchical structure among the data treating all relationships between variables at hierarchical level-1, or the individual level. This method of data analysis disregards between-group variance differences. Oppositely, aggregation is a simple linear regression method that disregards lower-level individual differences, instead of ignoring upper-level differences, such as disaggregation. Level-1 variables are treated at a higher level making variability among individuals disappear and all students are treated as homogenous entities (Gill, 2003; Woltman et al., 2012). Byrk and Raudenbush (1992) reported approximately 80%-90% of deviation among individual differences is lost when the aggregation method is used. HLM is often chosen instead of aggregation about hierarchical information due to its successful separation of the group and individual effects on the variable of interest (Woltman et al., 2012).

In HLM, mathematical theory and equations help conceptualize the concept of lower-level units representing individual students and higher-level units representing classrooms or whole grade levels. The complexity of HLM calculations increases exponentially with every increase in hierarchical level (Raudenbush & Bryk, 2002, Woltman et al., 2012). Therefore, the mathematical equations were more simply described based on a two-level hierarchical model. Of note, a two-level model was utilized in this research study instead of a three-level model due to these reasons. The simple linear regression created for each student *i:*

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \quad (1)$$

where:

Y_{ij} = dependent variable for the *i*th level-1 unit nested in the *j*th level-2 unit,

X_{ij} = value on the level-1 predictor,

$_{0j}$ = intercept for the *j*th level-2 unit,

$_{1j}$= regression coefficient associated with the X_{ij} for the *j*th level-2 unit, and

r_{ij} = random error associated with the *i*th level-1 unit nested within the *j*th level-2 unit.

(Woltman et al., 2012)

One important assumption of HLM is that any level-1 errors (r_{ij}) are normally distributed with a mean of zero and a variance equal to σ^2 (Woltman et al., 2012). For a level-2 model, the level-1 regression coefficients (β_{0j} + β_{1j}) are used as outcome variables and are connected to the level-2 predictors. The mathematical equation at level-2 becomes increasingly complex and supports the use of computerized statistical modeling programs. Level-2 models are also called between-unit models as they describe the variability among many groups (Gill, 2003; Woltman et al., 2012). Equations 2 and 3 conceptualize the case of a single level-2 predictor:

$$\beta_{0j} = {}_{00} + {}_{01}G_j + U_{0j} \qquad (2)$$

$$\beta_{1j} = {}_{10} + {}_{11}G_j + U_{1j} \qquad (3)$$

where:

β_{0j} = intercept of the *j*th level-2 unit;

β_{1j} = slope for the *j*th level-2 unit;

G_j = value on the level-2 predictor;

${}_{00}$ = overall mean intercept adjusted for G;

${}_{01}$ = regression coefficient associated with G relative to level-1 intercept;

${}_{11}$ = regression coefficient associated with G relative to level-1 slope;

U_{0j} = random effects of the *j*th level-2 unit adjusted for G on the intercept;

U_{1j} = random effects of the *j*th level-2 unit adjusted for G on the slope.

It is important to note that the level-2 (classroom) model brings two new terms (U_{0j} and U_{1j}) that Woltamn et al. (2012) identified as both unique to HLM. This allows for the model to determine an estimation of error that normal linear regression cannot determine. The covariance between β_{0j}, the intercept of the *j*th classroom, and β_{1j}, the slope for the *j*th classroom is equal to the covariance between U_{0j}, the random effects of the *j*th classroom, and U_{1j}, the random effects of the *j*th classroom adjusted on the slope. The assumptions of the level-2 (classroom) models are as follows (Raudenbush & Bryk, 2002, Woltman et al., 2012):

$$E(U_{0j}) = 0; \mathrm{E}(U_{1j}) = 0$$

$$\mathrm{E}(\beta_{0j}) = {}_{00}\ ; E(\beta_{1j}) = {}_{01}$$

$$\mathrm{var}(\beta_{0j}) = \mathrm{var}\ (U_{0j}) = {}_{00}\ ; var(\beta_{1j}) = var(U_{1j})\ {}_{11} \qquad (4)$$

$$\mathrm{cov}(\beta_{0j}, \beta_{0j}) = cov(U_{0j}, U_{1j}) = {}_{01}$$

$$cov(U_{0j}, r_{1j}) = cov(U_{1j}, r_{1j}) = 0$$

A combined model of equations 2 and 3 can be created to create equation 5. This combined model contains both level-1 (student) and level-2 (classroom) predictors and a term going across levels. The error is represented by $U_{1j}X_{ij} + U_{0j} + r_{ij}$. Equation 5 is called a mixed model as it contains both random and fixed effects which are unique to HLM (Woltman et al., 2012):

$$Y_{ij} = {}_{00} + {}_{10}X_{ij} + {}_{01}G_j + {}_{11}G_jX_{ij} + U_{1j}X_{ij} + U_{0j} + r_{ij} \qquad (5)$$

The combined model in equation 5 is similar to a normal regression but adds two new terms, U_{1j} and U_{0j}, which introduces an error estimation that is not a part of normal regression. Equation 5 demonstrates a dependency between level-1 (student) clustered within level-2 (classroom). In addition, U_{1j} and U_{0j}, may possess different values between level-2 units creating heterogeneous variances of the error terms.

IBM SPSS software originally stood for Statistical Package for Social Sciences but now is only known by the acronym. This widely-used statistical analysis package was used to perform HLM statistical calculations in this research study. The IBM SPSS data software was used due to its ability for prediction among nested groups using cluster analysis of HLM.

There are several advantages to using the HLM. First, the ability of HLM to simultaneously determine relationships between groups makes it very efficient for determining variances among different leveled groups. Second, HLM does not violate any of the statistical assumptions necessary in using older statistical techniques such as disaggregation and aggregation. This type of multilevel modeling is more accepting of violations in observation independence, homogeneity, and sphericity. HLM has little effect on standard errors, effect size, and variances. Finally, HLM is an effective method of determining the differences in variances between nested data as discussed previously (Gill, 2003; Woltman et al., 2012).

Although there are many advantages to using HLM and other multilevel, statistical modeling techniques, Dedrick et al. (2009) discussed several limitations and concerns to using HLM for data analysis that apply to this research study. The limitations fall into one of four categories: model development, hypothesis testing, data considerations, and estimation processing. Model development and specification issues are troublesome when determining and selecting predictor variables. The regression equations can become extremely complicated, especially when it is not feasible for predictor variables to have zero points. A second limitation related to using HLM for hypothesis testing and statistical inference occurs when sample sizes or variances parameters are small. Due to limited sample size, degrees of freedom may need adjustment when there is a violation of normality. Dedrick et al. (2009) commented that there are two methods of making inferences to overcome this problem. One method is to estimate the level-1 (grade level) coefficient separately from level-2 (classroom level) using OLS, which has its limitations. Also, the researcher can use empirical Bayes estimates, which consider all data but bias estimates. However, Bayes estimates tend to generate values more accurately than the parameter values. Dedrick et al. (2009) stated that there is, in fact, no estimation method that satisfies all conditions. Considerations of sample size and normality can assist researchers in determining which estimation method is the most appropriate for the research study (Dedrick et al., 2009).

An additional limitation to using HLM emphasized by Woltman et al. (2012) for traditional applications of HLM, is that substantial sample sizes are necessary at each level for sufficient power. Recently, researchers have overcome this problem by increasing the number of groups instead of increasing the number of observations per group. Groups of less than 50 could yield biased approximations of standard errors at the second level (Woltman et al., 2012).

Determining the Impact of STEM Education

There are three research studies that sought to determine the impact of STEM education in recent literature similar but with variation to this work. First, Wade-Shepard (2016) investigated the effect of middle school STEM curriculum on both science and math achievement scores using two ANCOVAs and the Pearson correlation to determine the strength of the relationship between STEM curriculum and student performance. Wade-Shepard (2016) did not include data analysis of other moderators and found a significant, strong, and positive correlation between test scores of students participating in STEM classes compared to those that were not taking STEM classes. Second, Hansen and Gonzalez (2014) investigated the relationships between STEM learning principles and student achievement in math and science. This mixed methods study found 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). Last, Han et al. (2015) analyzed the impact of STEM programming on math performance only using HLM among low, middle, and high achieving students accounting for moderators of student race and socioeconomic status. 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 socioeconomic status were strong predictors of student academic achievement (Han et al., 2015). This investigation is a variation of these three studies and seeks to determine the impact of STEM programming using HLM accounting for both math and science using additional moderators of gender, socioeconomic status, student race, and attendance rate.

Method

Participants

The participants were all students in grades 5-8 enrolled in the district and teaching staff from 2012-2013 to 2018-2019. Approximately 25-28 students were 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.

Setting

The setting for the research investigation was a mid-sized urban district located in northeastern Ohio. There were 1,061 students enrolled in the middle school serving grades 6-8 in the 2018-2019 academic year. (ODE, 2019). Both the percent of students in the district economically disadvantaged and the minority enrollment have remained steady over the seven years. The percent of students deemed economically disadvantaged, as defined by the number of students receiving a free or reduced lunch, ranged from 51-54%. The racial composition of the district was 81% White (non-Hispanic), 10% Black (non-Hispanic), 3% Multiracial, 3% Hispanic, and less than 1% Asian or Pacific Islander. By 2019, the number of students identified as economically disadvantaged had remained steady at 51%, with a student enrollment distribution similar to the beginning of the study of 75% White (non-Hispanic), 13% Black (non-Hispanic), 6% Multiracial, 6% Hispanic, and less than 1% Asian or Pacific Islander (ODE, 2019).

Instrument

The student measure of academic achievement was determined using annual Ohio State Test (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 in their 7th and 8th-grade year and science in 8th grade. The reliability and validity of Ohio state assessment data are reported annually through the ODE. Reliability of all OSTs across subjects and test years 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). The validity of using OST scores as an instrument for determining achievement is commonly used by researchers and is considered one of the stronger methods of defining achievement. State assessments are norm-referenced and standardized to ensure alignment with Ohio's Learning Standards for each grade level and subject. There was the suspension of state testing for the 2019-2020 school year due to the COVID-19 viral outbreak which caused school closure and suspended state testing. Therefore, no student test data were collected for the 2019-2020 academic school year.

The goal of this study was to determine the academic achievement of students participating in an integrated STEM program compared to students receiving traditional education. This longitudinal study commenced in the academic year 2012-2013 and concluded in 2018-2019. Students receiving STEM programming in 7th and 8th grade were considered the treatment group with general education students the control group. This was a two-level HLM analysis with level-2, school-related variables about grade levels five through eight depending on the school year, and the other variable of STEM participation. Level-1 variables are located within level-2 groups. The level-1, student-related variables consisted of OST scores, gender, race, socioeconomic status, attendance rate.

In context to the research problem the variables can be redefined as follows:

Y_{ij} = OST scores for student *i* in classroom *j*

X_{ij} = Participation in a STEM Program for student *i* in classroom *j*

$_{0j}$ = OST scores for student *i* in classroom *j* who did not participate in a STEM program

$_{1j}$ = regression coefficient associated with participation in a STEM program for the *j* classroom

r_{ij} = random error associated with student *i*

In context to the research problem, the variables can be redefined as follows when considering any level-1 predictor variable, in this example socioeconomic status:

β_{0j} = intercept of the *j*th classroom;

β_{1j} = slope for the *j*th classroom;

G_j = participation in a STEM or general education program;

${}_{00}$ = overall mean intercept adjusted for SES;

${}_{10}$ = overall mean intercept adjusted for SES;

${}_{01}$ = regression coefficient associated with SES relative to level-2 intercept;

${}_{11}$ = regression coefficient associated with SES relative to level-2 slope;

U_{0j} = random effects of the *j*th classroom adjusted for SES on the intercept;

U_{1j} = random effects of the *j*th classroom adjusted for SES on the slope.

The example given above illustrates how to determine the effect of variances between groups among hierarchical data. Socioeconomic status (SES) is one of the level-1 (student) predictors that was analyzed using HLM to determine its effect on students' OST scores as a dependent function of a level-2 (classroom) predictor of either STEM or general education programming.

According to Woltman et al. (2012), five conditions must be met to use HLM appropriately and effectively. Conditions two and three must be met before conditions four and five. The following conditions apply to this research study:

Condition 1: There is systematic within- and between-group variance in OST scores.

Conditions 2 and 3: There is significant variance in the level-1 (student) intercept and slope.

Condition 4: The variance in the level-1 (student) intercept is predicted by participation in a STEM or general education program.

Condition 5: The variance in the level-1 (slope) is predicted by participation in a STEM or general education program.

For this research design, the level-2 (grade-level tests) had approximately 9 groups and level-1 (classroom level) contained approximately 3,000-4,000 students, depending on the test year. There are a few limitations to the data collection and research design not necessarily related to the use of HLM. The study population is limited to a single institution versus data collection from other similar programs in the state. If other programs were included, the significance of the individual program teachers and instructors would be diminished. In this dataset, the race of the student was self-reported by the parent or guardian, leading to a degree of potential inaccuracy. Other factors seemingly unrelated to the STEM program may have effects on student performance and achievement. It is hypothesized that teacher experience and self-efficacy play an important role in student achievement and intellectual development. Finally, migration of students in and out of the STEM program may be a threat that affects both internal validity and reliability.

Table 1 displays the two hierarchical levels defining their category and factors, also known as variables at each particular level. HLM was used to analyze OST data for students in grades fifth through eighth to determine the effects of student achievement, the outcome variable, as a function of varying hierarchical levels.

Table 1. Variables at Each Hierarchical Level

Hierarchical Level	Category	Variables	HLM Variable Code
Level-2	School Level	Participation in a STEM program	STEMMARK
		Assessment Type	ASSMTTYP
		5th and 8th Grade Science tests	GRADE
Level-1	Student Level	OST scaled score	SCALEDSC
		Gender	GENDER
		Race	RACE
		Socioeconomic Status	SES
		Attendance	ATTEND

Variable and Sample Descriptions

Level 1 variables and corresponding coding are shown in Table 2. An economic disadvantage was coded=1 with not economically advantaged= 0. Gender was coded as female=1 and male= 0. Student race was coded as White (Non-Hispanic)=7, Puerto Rican= 6, Multiracial=5, Hispanic=4, Black (Non-Hispanic)=3, Asian=2, Alaskan Native/ Am. Indian=1. Attendance rate was coded as percent attendance at a decimal rate. This was calculated for students by taking the number of present days for the specific school year and dividing by the summation of days present, days absent (unexcused), and days absent (excused). The quotient was a decimal rate used as the ATTEND code for a given student in a specific school year.

Table 2. Student Level Variables with Label and Variable Coding

Student Level Variables	Label	Variable Coding
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.
OST Score	[Year_Subject]	Ohio State Assessment scaled score for a given academic year and subject; Subject is coded as Math or Science; Year is 2013 through 2019; 5th grade Science data 2010 through 2016

Table 3 displays the school-level variables with the given IBM SPSS (V. 26) labels and codes indicating student grade levels of five, seven, and eight. Student participation in integrated STEM programming was indicated by 1 and student participation in a general education setting was shown by 0.

Table 3. School Level Variables with Label and Variable Coding

School Level Variables	Label	Variable Coding
Grade	Grade	Grade 5= 5, Grade 7=7, Grade 8=8
STEM participation	STEM	STEM participation=1, General education participation=0

Missing Data and Outliers

There was a substantial number of participants missing attendance information for any given year. In SPSS, those cells were identified as missing in SPSS using ATTEND=9999 and coding "9999" as a missing data point. In addition, there were students missing OST scores. Those students were removed from the study. There were no outliers indicated in the data.

Descriptive Statistics

Descriptive statistics for student OST scores are segregated by STEM participation for each year of the study indicating the mean, standard deviation, maximum, minimum, skewness, and kurtosis. For every test, the mean OST score for students participating in the STEM program was higher than the mean score of general education students. Descriptive statistics were calculated for gender, student race, socioeconomic status, and attendance rate between the STEM program students and general education students to assist with determining interaction effects, if any.

Inter-correlations Between Variables

Preliminary bivariate relationships between STEM participation and student-level variables were determined using SPSS (V. 26) showing the Pearson correlation coefficients between student integrated STEM participation and the student-level coded variables. Out of the 3237 and 3240 students with data for GENDER and RACE, the Pearson bivariate correlation was extremely low (r= 0.069 and -.049, respectively) indicating a lack of practical significance. The *r* values, although statistically significant, are inflated due to the very large sample size of participants in the study. Therefore, there is no practical significance between these student-level variables and STEM participation.

The Pearson correlation coefficients between student integrated STEM participation and the student-level coded variable of SES by school year. For every school year, the number of students with socioeconomic information ranged from 405 for the 2011-20112 year to 1253 for the 2012-2013 school year due to the number of participants with OST data. The bivariate correlation was extremely low and lacked practical significance every year ranging from -.08 to 0.02. This indicates a lack of correlation between the student-level variable of socioeconomic status and STEM participation.

The Pearson correlation coefficient between student integrated STEM participation and the student-level coded variable ATTEND. For every school year, the number of students with attendance data ranged from 405 for the 2011-2012 year to 1253 for the 2012-2013 school year due to the number of participants with OST data. The bivariate correlation was low and lacked practical significance every year ranging from -.065 to 0.102. This indicates no practical correlation to a very small correlation between the student-level variable of attendance rate and STEM participation for the study years.

Preliminary bivariate relationships between STEM participation and student-level variables of gender, race, socioeconomic status, and attendance were all determined to be extremely low indicating no significant relationship and correlation between these student-level variables and STEM participation. Therefore, no HLM analysis was conducted using any of these student-level variables.

Hierarchical Linear Modeling Related to Case Study

The primary research question addressed in the case study was, "Does middle school integrated STEM programming positively affect student achievement"? This question was broken down into two sub questions answered by two HLM models.

Sub question 1: Does middle school integrated STEM programming positively affect student achievement in both math and science combined? (Model 1)

Sub question 2: Does middle school integrated STEM programming positively affect student achievement in science? (Model 2)

Sub Question 1: Does middle school integrated STEM programming positively affect student achievement in both math and science combined?

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 descriptive statistics for level-1. There were 8874 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.

Table 4. Model 1: Level-1 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
STEMMARK	8874	0.06	0.24	0.00	1.00
SCALEDSC	8874	598.94	149.06	314.00	868.00

Table 5 displays the descriptive statistics at level-2. There are nine tested years (N= 9) combining both math and science OSTs.

Table 5. Model 1: Level-2 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
ASSMTTYP	9	1.22	0.44	1.00	2.00

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

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

Level-2 Model

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

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

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 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. The proposed mixed model was found to be significant in predicting student achievement as a function of the defined level-1 and level-2 variables as shown below in Table 33 and Table 34.

Mixed Model

$$SCALEDSC_{ij} = \gamma_{00} + \gamma_{01}*ASSMTTYP_j + \gamma_{10}*STEMMARK_{ij} + u_{0j} + r_{ij} \qquad (8)$$

The final estimation of fixed effects with robust standard errors for Model 1 is shown in Table 6. 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 is predicted to score 31.3 points higher than general education students. All p-values are very small ($p \leqq 0.004$) supporting the correlation between STEM program participation and student achievement.

Table 6. 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

The final estimation of variance shown in Table 7 displays the random error associated with the use of the final estimation of fixed effects.

Table 7. 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			

Sub Question 2: Do students participating in an integrated middle school STEM program demonstrate differences in academic achievement in science?

Model 2- Comparing 5th and 8th Grade 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 8 displays descriptive statistics for both level-1. There were 4048 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.

Table 8. Model 1: Level-1 Descriptive Statistics

Variable Name	N	Mean	SD	Min.	Max.
STEMMARK	4048	0.06	0.24	0.00	1.00
SCALEDSC	4048	562.50	156.94	314.00	868.00

Table 9 displays the level-2 descriptive statistics. There were nine science assessments given with a minimum at 5th-grade and a maximum at 8th-grade.

Table 9. Model 2: Level-2 Descriptive Statistics

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

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

$$SCALEDSC_{ij} = \beta_{0j} + \beta_{1j}*(STEMMARK_{ij}) + r_{ij} \qquad (9)$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(GRADE_j) + u_{0j} \qquad (10)$$
$$\beta_{1j} = \gamma_{10}$$

The mixed model below (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 as shown below in Table 37 and Table 38.

Mixed Model

$$SCALEDSC_{ij} = \gamma_{00} + \gamma_{01}*GRADE_j + \gamma_{10}*STEMMARK_{ij} + u_{0j} + r_{ij} \qquad (11)$$

The final estimation of fixed effects with robust standard errors for Model 2 is shown in Table 10. 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 very small (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.

Table 10. 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 final estimation of variance shown in Table 11 displays the random error associated with the use of the final estimation of fixed effects.

Table 11. 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			

Discussion

The current investigation examines the impact of student participation in a STEM education program on student achievement as measured by performance on state level mathematics and science assessments. This research includes multiple years of data. HLM is uniquely prepared to deal with this type of analysis, as it provides the researcher with the ability to look within the natural setting with a thorough analysis that can dissect what is occurring at various levels in the data (Woltman et al., 2012). Through HLM, multiple linear levels of data as well as nested data, such as student within a particular school year, can be simultaneously analyzed, thus maximizing the power of the analysis while accounting for potential auto-correlation (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002). HLM provides the ability to understand where effects may be occurring in educational data which rarely includes any form of random assignment. As such HLM provides a rigorous analysis with good external validity that can provide educators and researchers with valuable insights (Shadish et al., 2001; Trochim et al., 2016). Additionally, the current investigation benefits from the use of multiple years of data, which increases the reliability and confidence in the computed estimates of the overall impact (Shadish et al., 2001).

Model 1 used OST score and STEM participation at level-1 and the assessment type by year at level-2. Descriptive statistics indicated all participants at level-1 (n=8874) had a mean OST score of 598. 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. 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 very small ($p \leqq 0.004$) supporting the correlation between STEM program participation and student achievement.

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. Descriptive statistics indicated all participants at level-1 (n=4048) had a mean OST score of 562 points. 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 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 very small ($p \leqq 0.009$) supporting the correlation between STEM program participation and student achievement in science except for 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 in Model 2 compared to both math and science achievement in Model 1. The use of HLM is unique in its ability to enable these outcomes to be discovered within this multi-year and multi-layered data set (Bryk & Raudenbush, 1992; Raudenbush & Bryk, 2002).

There is no single statistic that can fully quantify the success of STEM education on student performance (Gonzalez & Kuenzi, 2012). However, the use of HLM to determine the impact of STEM education in nested datasets, such as classrooms within schools across years is a powerful tool to quantify such effects (Raudenbush

& Bryk, 2002). This investigation was an amalgam of three studies previously discussed: Wade-Shepard (2016); Hansen & Gonzalez (2014); Han et al. (2015). The results of this investigation indicate through comparison of descriptive statistics and HLM analysis that middle school students participating in integrated STEM programming scored significantly higher on the OST compared to their general education peers. This aligns with earlier findings by Wade-Shepard (2016) and Hansen and Gonzalez (2014) using different methods of analysis. Although student race and socioeconomic status are often correlated with student growth and achievement similar to what was reported by Han et al. (2015), the present investigation did not conclude similar results to support the claims in previous literature.

Conclusion

Hierarchical linear modeling has become an increasingly popular multilevel method of analyzing data among nested datasets, in particular, the effect of specialized programming within schools. The purpose of this methodological study is to demonstrate the use of HLM to determine the effectiveness of STEM programming in an Ohio middle school. This longitudinal study used student effects of gender, socioeconomic status, student race, and attendance rate along with state test scores to quantify achievement. HLM analysis determined STEM programming had a significant, positive effect on achievement in math and science with no interaction effects determined between STEM participation and student factors. This demonstrates HLM as a powerful statistical tool in analyzing the impact of specialized academic programs on student achievement within nested datasets, such as classrooms within schools.

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Sage.

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., ... & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79*(1), 69-102.

Dugger, W. E. (2010, December). Evolution of STEM in the United States. In *6th biennial international conference on technology education research* (Vol. 10).

Gill, J. (2003). Hierarchical linear models. *Encyclopedia of Social Measurement*. Elsevier.

Gonzalez, H. B., & Kuenzi, J. J. (2012, August). Science, technology, engineering, and mathematics (STEM) education: A primer. Congressional Research Service, Library of Congress.

Han, S., Capraro, R., & Capraro, M. M. (2015). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement. *International Journal of Science and Mathematics Education, 13*(5), 1089-1113.

Hansen, M. & Gonzalez, T. (2014). Investigating the relationships between STEM learning principles and student achievement in math and science. *American Journal of Education. 120*(2). 139-171.

Ohio Department of Education. (2014, July). Ohio achievement assessments statistical summary.

https://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

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Snyder, M. (2018). A century of perspectives that influenced the consideration of technology as a critical component of STEM education in the United States. *The Journal of Technology Studies*, *44*(2), 42-57.

Trochim, W. M., Arora, K., & Donnelly, J. P. (2016). Research methods: The essential knowledge base. Cengage Learning.

Wade-Shepard, A. A. (2016). *The effect of middle school STEM curriculum on science and math achievement scores.* [Doctoral dissertation, Union University].

White, D. W. (2014). What is STEM education and why is it important? *Florida Association of Teacher Educators Journal, 1*(14), 1-9.

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.

Author Information

Danielle R. Chine
https://orcid.org/0000-0001-8196-3252
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
United States
Contact e-mail: *dchine@andrew.cmu.edu*

Karen H. Larwin
https://orcid.org/0000-0002-7574-6291
Youngstown State University
One University Circle
Youngstown, OH 44555
United States