Journal of Education and Human Development
December 2016, Vol. 5, No. 4, pp. 98-108
ISSN: 2334-296X (Print), 2334-2978 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jehd.v5n4a10
URL: https://doi.org/10.15640/jehd.v5n4a10

The Impact of Flipped Instruction on Middle School Mathematics Achievement

Amanda Martin, Ed.D[1], Melissa Arrambide, Ed.D.[2] & Chuck Holt, Ed.D.[3]

Abstract

The purpose of this causal comparative quantitative research study was to examine the effectiveness of flipped instruction on middle school mathematics achievement. The effectiveness of the flipped classroom in closing the existing achievement gap among students of various ethnic sub-populations, socio-economic statuses, and within pre-AP mathematics classes was investigated in this study. Propensity score matching was used to match students taught by the same teacher within control and treatment groups using 1:1 nearest neighbor matching with a caliper of 0.25 *SD*. The matched student data were analyzed using multilevel modeling. The results indicated that no significant differences existed between the STAAR Mathematics scale scores of African American, Hispanic, White, economically disadvantaged, non-economically disadvantaged, and pre-AP mathematics students in flipped or traditional classrooms.

Time is a valuable resource, and strategies are needed to increase learning and teaching time. Educational leaders can impact student achievement through the selection and implementation of effective instructional strategies (Onorato, 2013). Flipped instruction, or the flipped classroom, is an instructional approach that allows educators to "radically rethink how they use class time" (Tucker, 2012, p. 82). It is gaining momentum and attention in the learning community and could be selected by educational leaders to positively influence student achievement.

The flipped classroom essentially 'flips' what is traditionally done in class and what is traditionally done as homework. The flipped classroom is an "inverted approach in which the students' homework is to view a recording of the lecture, and class time is used for active problem-solving activities with instructor guidance" (DeMaio & Oakes, 2014, p. 340). In the traditional mathematics classroom, students listen to the instructional lecture during class time and practice mathematics problems in the time remaining at school and finish at home. In contrast, in the flipped mathematics classroom, students view instructional lecture videos at home and practice mathematics problems during class time. The flipped classroom could offer a solution for teachers and administrators who want to maximize the use of class time to increase student achievement (DeMaio & Oakes, 2014).

In the flipped classroom "students learn by doing, and ... the doing is happening within a hand-raise of the teacher. Students are no longer at home in isolation and unsupported while they do the difficult work of learning" (Fulton, 2012, p. 22). When students watch lecture videos prepared by their teacher outside of class time, they are prepared to engage in meaningful practice during class. This also allows the teacher to differentiate for students and work with individual students or small groups each day. Differentiated instruction points to "achievement gains on standardized tests, including mathematics assessments" (Chamberlin & Powers, 2010, p. 116).

[1] Mathematics Facilitator, Mesquite Independent School District. amandamartin831@gmail.com
[2] Assistant Professor, Master's Degree and Principal Preparation Program Coordinator, Texas A & M University Commerce, Department of Educational Leadership, USA. Melissa.Arrambide@tamuc.edu, 214-405-5619
[3] Assistant Professor, Interim Department Head, Texas A & M University Commerce, Department of Educational Leadership, USA. Chuck.Holt@tamuc.edu, 903-353-4112

Teachers can engage students in learning activities tailored to their needs and "most importantly, all aspects of instruction can be rethought to best maximize the scarcest learning resource – time" (Tucker, 2012, p. 82). Time for practice and solving problems is valuable, and using technology maximizes learning time in the flipped classroom.

Significance of the Research

The intent of the causal comparative quantitative study was to identify whether the flipped instructional strategy was an effective approach to increase mathematics achievement for middle school students. Closing the achievement gap among ethnic sub-populations is a goal for many academic programs (Chamberlin & Powers, 2010). African American and Hispanic students are expected to exemplify success on state assessments at the same level as their White peers, and an instructional approach must be found to close the achievement gap. Mathematics achievement trends "suggest that the gap between some minority and White students persists and may even be widening" (Bol & Berry, 2005, p. 33). Student achievement is paramount, and educational leaders seek effective instructional strategies to employ in the learning environment to close the achievement gap among students of various ethnic sub-populations. In addition, studies reveal that economically disadvantaged students typically perform below their mainstream counterparts and "there is a considerable gap in test performance between students from poor families and those from non-poor families" (Flores, 2007, p. 30). Teachers and administrators need instructional approaches that positively impact student achievement and creatively use valuable instructional time.

Statement of the Problem

The accountability system in Texas is rigorous and holds districts, administrators, teachers, and students to high standards. Students in the state of Texas are evaluated annually for performance on the State of Texas Assessment of Academic Readiness (STAAR). School districts in Texas must ensure that all students learn required content and must look carefully at achievement gaps between students of various ethnicities and socio-economic statuses (Texas Education Agency, 2010). In the state of Texas, 79% of students in all grades met expectations on the 2013 State of Texas Assessment of Academic Readiness (STAAR) Mathematics assessment. However, an achievement gap exists between ethnic sub-populations. White students outperform their African American and Hispanic peers. In addition, students who are economically disadvantaged perform slightly lower than non-economically disadvantaged peers. Efforts must be made to close the achievement gap. Teachers are looking for innovative ways "to instruct technology age students, administrators are seeking new ways to lead teachers in an age of increasingly uncertain resource allocations, and district officers are looking for new ways to train instructional leaders for the 21st century" (Smith & Addison, 2013, p. 135). As new instructional strategies are needed to close the achievement gap and engage students, flipped instruction was examined.

Purpose of the Study

The purpose of this quantitative study was to examine the effects of the flipped classroom on middle school mathematics achievement when compared to the traditional classroom. Another aim was to determine the effectiveness of the flipped classroom as an instructional approach in closing the existing achievement gap among identified student groups, including African American, Hispanic, White, and economically disadvantaged and non-economically disadvantaged. The effectiveness of the flipped classroom in the instruction of students within pre-AP classrooms was also examined. To measure student achievement, scale scores from the STAAR Grade 8 Mathematics and STAAR Grade 7 Mathematics, used as a covariate, were reported, collected, and analyzed. Flipped instruction is a new instructional strategy used in the classroom with little research regarding its effectiveness. The results of the study can inform educational leaders as they choose effective instructional strategies to implement in the learning environment.

Method of Procedure

This causal-comparative quantitative study analyzed the effectiveness of the flipped instructional strategy on mathematics achievement of Grade 8 students when compared to the traditional instructional approach. The quantitative data of this study included the STAAR Grade 8 Mathematics scale scores, student ethnicities, socio-economic status, and level of instruction, regular education or pre-AP. The independent variable was defined as the type of instruction, flipped instruction or traditional instruction.

The dependent variable was mathematics achievement as measured by the STAAR Grade 8 Mathematics scale score, using the STAAR Grade 7 Mathematics scale score as a covariate. The scale score is "a conversion of the raw score onto a scale that is common to all test forms for that assessment. Scale scores allow for direct comparisons of student performance between specific sets of test questions from different test administrations" (Texas Education Agency, 2012). The scale score accounts for the difficulty level of the assessment and quantifies the performance of the student (Texas Education Agency, 2012). The students receiving instruction in the traditional classroom served as the control group, and the students receiving instruction in the flipped classroom served as the treatment group.

Selection of Sample

The causal-comparative quantitative study was conducted in one suburban independent school district in Texas. The district served approximately 39,000 students in grades Pre-Kindergarten through 12. The student ethnic make-up of this district included 51.3% Hispanic students, 24.9% African American students, 19.3% White students, and 4.5% other students (Texas Education Agency, 2013a). Currently, 70.3% of students in the suburban district are considered economically disadvantaged (Texas Education Agency, 2013a). This study included the data from all 8 middle school campuses. In the suburban independent school district used in this study, 75% of students met expectations on the 2013 State of Texas Assessment of Academic Readiness (STAAR) Grade 7 Mathematics. An achievement gap existed between ethnic sub-populations. Similar to the state, White students outperform their African American and Hispanic peers in the suburban independent school district used in this study. In addition, students who are economically disadvantaged performed lower than non-economically disadvantaged peers.

The district instructed Grade 8 students in regular mathematics classes, pre-Advanced Placement (pre-AP) mathematics classes, gifted and talented (GT) mathematics classes, and basic mathematics classes. The Grade 8 students who received grade-level instruction in regular and pre-AP mathematics classes were the target population for this study. Permission to access the suburban independent school district was obtained by the researcher and the data were retrieved by district appointed personnel.

Data Set

In this study, mathematics achievement was measured with the STAAR Grade 8 Mathematics scale score and the STAAR Grade 7 Mathematics scale score as the covariate. The control group consisted of the 2012-2013 Grade 8 students who received traditional instruction, and the treatment group consisted of the 2013-2014 Grade 8 students who received flipped instruction taught by the same teacher. It was important for each student included in the data set to have both a STAAR Grade 8 Mathematics scale score and a STAAR Grade 7 Mathematics scale score. Students without both scores were deleted from the data set.

The suburban independent school district conducted the Middle School Mathematics Flipped Instruction Survey in May 2014 to determine the Grade 8 mathematics teachers that utilized traditional and flipped instruction for the duration of the 2013-2014 academic years. Permission to access the survey results was granted and district appointed personnel retrieved the data. Through the use of cluster sampling, the student data of Grade 8 teachers who implemented flipped instruction for the duration of the 2013-2014 academic years served as the treatment group. Once the flipped classroom teachers were identified, their student data from the previous 2012-2013 academic year served as the control group. This control group received traditional instruction for the 2012-2013 academic years.

The 2013-2014 treatment and 2012-2013 control group students taught by the same teacher were included in this study. The student data of Grade 8 teachers that chose to implement flipped instruction for a shorter period of time were not included in the study. Of the 31 Grade 8 mathematics teachers, 11 teachers chose to implement flipped instruction for the duration of the 2013-2014 academic years. The student data of teachers that chose to implement traditional instruction were deleted from the data set as well as the teachers that chose to implement flipped instruction for a shorter time period. Of the 11 teachers that chose to implement flipped instruction for the duration of the academic year, 6 of the teachers also taught Grade 8 mathematics using traditional instruction during the 2012-2013 academic years. Therefore, the student data of the other 5 flipped classroom teachers that did not deliver traditional instruction during the 2012-2013 academic years were deleted from the data set. Once the data were collected for this study, the created data set included data from 1,025 students in the control and treatment groups.

The 2013-2014 treatment group included 388 students in regular mathematics classes and 115 students in pre-AP mathematics classes for a total of 503 students instructed in the flipped classroom setting. Taught by the same teachers, the 2012-2013 control group consisted of 383 students from regular classes and 139 students from pre-AP mathematics classes for a total of 522 traditional classroom students. The data set given to the researcher included a randomized student identification number and randomized teacher number. In addition, the student data included ethnicity, gender, socio-economic status, LEP status, special education status, level of instruction (regular or pre-AP), STAAR Grade 8 Mathematics scale score, and STAAR Grade 7 Mathematics scale score. The data were retrieved by district appointed personnel and delivered to the researcher in electronic spreadsheet form.

The researcher prepared the data for analysis. The ethnicity for each student was replaced with a code (1 = African American, 2 = Hispanic, 3 = White, 4 = other). The gender was also replaced with a code (0 = male, 1 = female). LEP status, special education status, and socio-economic status were replaced with a binary code (0 = no, 1 = yes). For level of instruction, a 1 represented a student who received instruction in a regular mathematics class and a 2 represented a student who received instruction in a pre-AP mathematics class. Each student taught by the same teacher received a class identification number as well. The STAAR Mathematics scale scores remained as reported.

Data Analyses

Once the data were prepared, the data were analyzed using version 21.0 of the Statistical Package for the Social Sciences (SPSS). The student data in the 2013-2014 treatment group file were disaggregated into regular and pre-AP mathematics classes. Descriptive statistics and frequencies were run for both levels of instruction for flipped classroom students. The student data in the 2012-2013 control group were also disaggregated into regular and pre-AP mathematics classes. Descriptive statistics and frequencies were analyzed for the traditional classroom students. This analysis was conducted to depict the student data before propensity score matching and was reported in Table 1 (regular mathematics students) and Table 2 (pre-AP mathematics students).

Propensity Score Matching

A propensity score matching (PSM) technique described by Thoemmes (2012) was used in this study to match students from the control and treatment groups. The statistical matching of students accounted for the non-randomization within the study because the students were matched across the treatment and control groups based on the covariates of ethnicity, socio-economic status, gender, LEP status, special education status, and scale scores for the STAAR Grade 7 Mathematics. Students were matched based on the propensity or probability of receiving flipped instruction (Stuart & Rubin, 2008). To further reduce teacher effect in this study, the students within the treatment and control groups received instruction from the same teacher. The propensity scores were calculated for each student in the study using 1:1 nearest neighbor matching, which means that “a single treated participant is matched to a single untreated participant who has the most similar estimated propensity score” (Thoemmes, 2012, p. 5). A caliper of 0.25 *SD*, also used by Rickles (2011), allowed for a close match between students of the flipped and traditional instruction groups. This propensity score matching technique allowed the researcher to analyze the differences in treatment and control groups although randomization and true experimental design was not possible. A propensity score matching procedure adapted from Randolph, Falbe, Manuel, and Balloun (2014) was used to match students.

The propensity score was estimated for each student within the control and treatment groups. First, the students who received instruction in regular mathematics were matched using a 1:1 nearest neighbor matching without replacement using a caliper of 0.25 *SD* with all control and treatment group students, matched exactly by teacher. The 279 matched pairs were saved as data set M. Differences on almost all covariates were smaller in the matched sample (Table 1). The greatest difference was in the pre-test STAAR Grade 7 Mathematics scale score, where in the matched sample, the traditional group had a mean score that was 9.2 points higher than the flipped sample.

Table 1: Student Characteristics of Grade 8 Regular Math Students Original and Matched Samples

Characteristic	Original Sample Traditional	Flipped	Difference	Matched Sample Traditional	Flipped	Difference
Number of Students	383.0	388.0	-5	279.0	279.0	0.0
% Female	48.8	46.9	1.9	49.5	45.2	4.3
Ethnicity						
% African American	34.5	35.8	-1.3	38.0	39.4	-1.4
% Hispanic	48.8	46.6	2.2	46.6	45.5	1.1
% White	13.8	12.9	0.9	14.0	13.3	0.7
% Econ. Disadvan.	69.2	67.0	2.2	65.6	67.7	-2.1
% LEP	9.7	6.4	3.3	5.0	4.3	0.7
% SpEd	6.8	5.7	1.1	5.0	5.0	0.0
Pre-Scale Score	1585.0	1577.8	7.2	1583.7	1574.6	9.2
Propensity Score	.49813	.50829	-0.01016	.50451	.50975	-0.00524

Data set MC was created by matching the control and treatment group students who received instruction in pre-AP mathematics classes using a 1:1 nearest neighbor matching without replacement using a caliper of 0.25 *SD*, matched exactly by teacher. The results of the propensity score matching results for the 77 matched student pairs in pre-AP mathematics classes are shown in Table 2. Differences on almost all covariates were smaller in the matched sample. The greatest difference was in the percent of economically disadvantaged students, where in the matched sample, the flipped group had 12.9% more economically disadvantaged students than the traditional sample.

Table 2: *Student Characteristics of Grade 8 Pre-AP Math Students Original and Matched Samples*

Characteristic	Original Sample Traditional	Flipped	Difference	Matched Sample Traditional	Flipped	Difference
Number of Students	139.0	115.0	24.0	77.0	77.0	0.0
% Female	59.0	53.9	5.1	62.3	62.3	0.0
Ethnicity						
% African American	20.9	23.5	-2.6	18.2	22.1	-3.9
% Hispanic	50.4	47.0	3.4	45.5	42.9	2.6
% White	19.4	21.7	-2.3	23.4	23.4	0.0
% Econ. Disadvan.	57.6	56.5	1.1	44.2	57.1	-12.9
% LEP	2.2	1.7	0.5	2.6	1.3	1.3
% SpEd	1.4	0.0	1.4	0.0	0.0	0.0
Pre-Scale Score	1705.3	1718.9	-13.6	1702.0	1706.2	-4.2
Propensity Score	.44322	.46429	-0.02107	.44493	.45137	-0.00644

Using propensity score matching, the matched pairs of student data were more similar to one another based on covariates. The covariates were the predictor variables (ethnicity, gender, socio-economic status, LEP status, special education status, STAAR Grade 7 Mathematics scale score), and the treatment status (0 = traditional, 1 = flipped) was the outcome variable. A logistic regression estimation algorithm was used, discarding units outside the common area of support. Selecting this option improved balance on covariates and eliminated the units when no comparable data were found (Thoemmes, 2012). Once the 279 regular mathematics student pairs and the 77 pre-AP mathematics student pairs were matched from control and treatment groups, standardized differences were computed. The standardized differences method recommended by Ho, Imai, King, and Stuart (2007) was used to assess the balance of the differences in this study as shown in Table 3. The standardized difference is "the absolute difference in sample means divided by an estimate of the pooled standard deviation of the variable" (Austin, 2008, p. 2039). This method does not depend on sample size and represented the difference in treatment and control group means expressed as units of standard deviation (Austin, 2008).

Characteristics of the treatment and control groups were computed for both the regular and pre-AP classes, and then the measure of the standardized difference between treatment and control classes, *d*, was calculated. For the regular classes, the greatest difference was in the pre-test STAAR Grade 7 Mathematics scale score (*d* = .114), however, this met the rule of thumb that the standardized difference be less than .25 (Austin, 2008). For the pre-AP students, the greatest difference was in the percent of economically disadvantaged students (*d* = .259). This was slightly higher than typically accepted, but because all other variables were balanced, this difference was accepted.

Table 3: Standardized Differences in Characteristics of Treatment and Control Groups by Level of Instruction, Regular or Pre-AP

Characteristic	Regular Treatment Mean	Control Mean	*d*	Pre-AP Treatment Mean	Control Mean	*d*
Ethnicity						
% African American	39.4	38.0	0.029	22.10	18.20	0.097
% Hispanic	45.5	46.6	-0.022	42.90	45.50	-0.052
% White	13.3	14.0	-0.021	23.40	23.40	0.000
% Econ. Disadvantaged	67.7	65.6	0.046	57.10	44.20	0.259
% Special Ed	5.0	5.0	0.000	0.00	0.00	-
% LEP	4.3	5.0	-0.034	1.30	2.60	-0.094
% Female	45.2	49.5	-0.086	62.30	62.30	0.000
Grade 7 Math STAAR Scale Score	1569.4	1574.6	-0.114	1706.17	1701.95	0.054

Restructuring the Data

Once the researcher had confirmed that the treatment and control groups were similar, Level 2 class variables for both Grade 7 and Grade 8 mathematics classes were merged into the matched student data set. These variables included the level of instruction (regular or pre-AP) and the percent of students in the class that were economically disadvantaged, female, LEP, and those that received special education services. Ethnicity dummy variables were also computed for the class data. Then the data set was restructured into a person-period format (Heck, Thomas, & Tabata, 2010) where each student had two lines of data, one for Grade 7 and one for Grade 8. New time-varying variables were computed to distinguish time (time 0 = Grade 7 pre-test scale score, time 1 = Grade 8 post-test scale score) and an index variable was created. The data were restructured to facilitate using the mixed model linear program in SPSS.

Multilevel Modeling

In this study, random assignment was not feasible. Therefore, "the combination of propensity score matching ... and multilevel modeling is a promising tool to examine causal effect heterogeneity in educational settings" (Rickles, 2011, p. 5). Multilevel modeling (MLM) was chosen for this study because of the nested nature of the educational data. "Multilevel modeling provides a powerful framework for analyzing data collected in the school context" (Dettmers, Trautwein, Ludtke, Kunter, & Baumert, 2010, p. 472). Multilevel modeling is a robust model of analysis and was used to analyze the matched pairs of student data. As in most research conducted "in school settings, students in this study are nested within classes. Students within a class are typically more similar to each other than are two students randomly selected from the whole sample" (Dettmers et al., 2010, p. 472). Multilevel modeling considered the nested data within the educational context. Level 1 student data were nested within Level 2 class data. The Level 1 student variables included the STAAR Grade 8 Mathematics scale score, gender, socio-economic status, ethnicity, special education status, LEP status, and the STAAR Grade 7 Mathematics scale score as the covariate. Level 2 class data included the type of instruction (flipped or traditional), level of instruction (regular or pre-AP), and aggregated class percent data. Because teachers and schools used either all flipped classes or no flipped classes, it was not necessary for teacher and school variables to be included in the model.

Findings and Implications

This causal-comparative quantitative study analyzed the impact of flipped instruction on the mathematics achievement of Grade 8 students in one suburban independent school district.

Data from students who received instruction in the flipped classroom and the traditional classroom were compared and analyzed. Results from the multilevel modeling of propensity score matched students (Table 4) were used to answer the research questions.

Table 4: Parameter Estimates for Five Models Examining the Differences Between Traditional and Flipped Instruction with Covariate Influence

Fixed Components	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1622.38***	1603.23***	1603.23***	1527.56***	1532.79***
Time1		68.27***	69.82***	70.00***	70.09***
Flipped			-3.09	-3.50	-41.05
Special Ed				-69.96***	-69.98***
Econ. Disadvantaged				-11.21	-11.37
Limited English Profic.				-49.01**	-48.80**
Gender				-15.21**	-15.25**
African American				-58.83***	-61.59***
Hispanic				-25.46	-29.03
White				-33.15*	-37.76*
Level of Instruction				111.01***	109.52***
Flipped*EconDis					0.35
Flipped*African Amer.					20.63
Flipped*Hispanic					25.89
Flipped*White					33.99
Flipped*Level of Instruc.					10.72
Variance of Random Components					
Repeated Measures					
Variance	2701.17***[a] 6928.93***[b]	1383.34***	1388.21***	1352.87***	1379.61***
Intercept + Time, Subject = Teacher					
Time	771.45[c]	55.00	54.93	49.24	54.62
Intercept		781.83	781.77	81.34	84.13
Intercept + Time, Subject = ClassID + Teacher					
Time	6171.71***[c]	2167.88***	2163.98***	2237.03***	2197.85***
Intercept		7119.71***	7114.84***	4390.27***	4363.20***
AIC	16996.199	16521.557	16516.082	16158.381	16121.382

Note. [a]Variance = Time 0. [b]Variance = Time 1. [c]Variance is not associated with Time.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Research Question 1

"To what extent does the flipped classroom increase student achievement on the STAAR Grade 8 Mathematics assessment when compared to the traditional classroom?" This research question compared the student data from the traditional and flipped classrooms. According to the null model (Table 4, Model 1), significant differences existed between scores over time from the STAAR Grade 7 Mathematics scale score to the STAAR Grade 8 Mathematics scale score ($p < .001$). Significant differences also existed within students by teacher ($p < .001$). However, no significant differences in scores were present within teachers (Wald $Z = 12.632$, $p = .161$). Additional analyses were needed to identify the variability among scores. The growth rate model (Table 4, Model 2) and treatment model (Table 4, Model 3) indicated significant differences in scores within students by teacher with a gain of approximately 68 points from the Grade 7 scale score to the Grade 8 scale score (Table 4, Model 2). When treatment was added to the multilevel model of the Level 1 student descriptors (Table 4, Model 4), the treatment of flipped instruction did not prove to increase student scores significantly ($\beta = -3.50$, $t = -.686$, $p = .493$).

Therefore, the researcher failed to reject the null hypothesis of no significant difference between the mathematics scores of students receiving instruction in the flipped classroom and the mathematics scores of students receiving instruction in the traditional classroom.

The null model did indicate an unusual finding. No significant differences were found in covariance estimates between teachers (Table 4, Model 1). This was unusual because the student data from six different flipped classroom teachers that previously used traditional instruction were analyzed in this study. While all teachers taught the same subject both years, some of these teachers were from different campuses within the same school district. This finding could be a result of the paced curriculum that the school district employed. All teachers within the suburban independent school district maintained consistent pacing and accessed the same curriculum resources and lesson documents. Common district assessments are used to monitor student progress and monitor the instructional program. Although it is unusual to not find differences between teachers (Level 2 groups), it does not change the interpretation of the fixed effects (Hox, 2010).

Research Question 2

"To what extent do the flipped classroom increase mathematics scores of economically disadvantaged students when compared to the traditional classroom?" This research question compared the data of economically disadvantaged and non-economically disadvantaged students from the traditional and flipped classrooms. The Level 1 model indicated that being economically disadvantaged was associated with an 11-point decrease in STAAR Mathematics scale scores although this was not a significant value (Table 4, Model 4). According to the interaction between the treatment of flipped instruction and the Level 1 student descriptor of economically disadvantaged (Table 4, Model 5), the scores of economically disadvantaged students in flipped and traditional classes did not show significant differences (β = .353, t = .044, p = .965). Therefore, the researcher failed to reject the null hypothesis of no significant difference between the mathematics scores of economically disadvantaged students receiving instruction in the flipped classroom and the mathematics scores of economically disadvantaged students receiving instruction in the traditional classroom.

Research Question 3

"To what extent does the flipped classroom close the achievement gap between students of ethnic sub-populations when compared to the traditional classroom?" This research question compared the student data of various ethnic sub-populations from the traditional and flipped classrooms. According to the interaction between the treatment and ethnicity, the beta estimates showed an increase in scores when the treatment was present (Table 4, Model 5). The decrease of approximately 59 points in the STAAR Grade 8 Mathematics scale score for African American students was statistically significant (Table 4, Model 4). However, although it was not statistically significant, the beta estimate for African American students in flipped classrooms gained about 21 points from the Grade 7 to Grade 8 STAAR scale score (β = 20.63, t = .1.024, p = .306). This finding showed that scores of African American students decreased significantly from Grade 7 to Grade 8.

However, the scores of African American students increased in the flipped classroom. This is a promising result. In addition, there was a decrease of 29 points from Grade 7 to Grade 8 for Hispanic students, but the decrease was not statistically significant (Table 4, Model 4). For Hispanic students, the beta estimate showed an approximate 26-point increase for those students in the flipped classes, although it was not significant (β = 25.89, t = .1.306, p = .192). White student's experienced significant losses of 33 points from Grade 7 to Grade 8 on the STAAR Mathematics scale score (Table 4, Model 4). However, for White students in flipped classrooms, the beta estimate showed an approximate 34-point increase for those students, but was not significant (β = 33.99, t = 1.621, p = .105). The ethnic sub-populations of African American, Hispanic, and White showed an increase in scores in the flipped classrooms, which revealed promising results. Given that the increases in scores were not statistically significant, the researcher failed to reject the null hypothesis of no significant difference between the mathematics scores of White, African American, and Hispanic students receive instruction in the flipped classroom and the mathematics scores of the White, African American, and Hispanic students receiving instruction in the traditional classroom.

Research Question 4

"To what extent do the flipped classroom increase mathematics scores of students in pre-AP and regular mathematics when compared to the traditional classroom?" This research question compared the student data in regular and pre-AP mathematics classes from the traditional and flipped classrooms. The model estimated a statistically significant 111-point increase for pre-AP students from Grade 7 to Grade 8 on the STAAR Mathematics scale score (Table 4, Model 4). Students within these courses are taught grade-level content with enrichment. Teachers exemplify high expectations for these students. According to the interaction of the flipped treatment with the level of instruction in regular and pre-AP mathematics classes (Table 4, Model 5), the beta estimate showed an approximate 11-point increase for those students in pre-AP classes using flipped instruction, although it was not a significant difference (β = 10.72, t = 1.147, p = .252). Therefore, the researcher failed to reject the null hypothesis of no significant difference between the scores of students in pre-AP and regular mathematics classes receiving instruction in the flipped classroom and the scores of students in pre-AP and regular mathematics classes receiving instruction in the traditional classroom.

Review of the Literature and Discussions

The review of the literature indicated positive attributes of the flipped classroom. The flipped classroom is gaining momentum in the learning community, yet little quantitative research exists on the effects of the flipped classroom (Love, Hodge, Grandgenett, & Swift, 2014). Existing survey research indicated that the flipped classroom improved teacher job satisfaction, increased test scores, improved student attitudes, and provided time to differentiate instruction (Brunsell & Horejsi, 2013). In addition, Chamberlin and Powers (2010) noted that studies in classes using differentiated instruction, like in flipped classrooms, had seen an impact on student achievement for all students and across all ethnic sub-populations and for students of various socio-economic statuses. In the context of this research study, the flipped classroom did not prove to have a significant impact on the achievement of all middle school mathematics students or for economically disadvantaged students. However, the ethnic sub-populations in this study experienced increased test scores in flipped classrooms, although the gains were not statistically significant (Table 4, Model 5).

Most of the existing research regarding the flipped classroom comes from courses in higher education. Baepler, Walker, & Driessen (2014) noted that the results of the flipped college chemistry course benefitted from the use of flipped instruction in that learning outcomes were either statistically equal to or superior to the outcomes in traditional classrooms. This was noteworthy because the traditional classroom met three times per week with lecture format but the flipped classroom was restructured to meet one time per week with instructional videos and online discussion. Millard (2012) noted that the flipped classroom adjusts the delivery style and learning responsibilities to the student and higher education courses are succeeding with flipped classrooms. This is true for college courses, but for this research study, the achievement of middle school students was analyzed. The age and maturity of the student may impact the effectiveness of the flipped classroom. In addition, a study about a flipped geometry course revealed that students reviewed the instructional lecture video multiple times to understand the content (Gullen & Zimmerman, 2013). While not measured in this study, the amount of effort and time spent studying content at home could have an impact on the effectiveness of the flipped classroom along with other student responsibilities.

The responsibilities of the teacher could also have an impact on the results of the flipped classroom's effectiveness. Thiele (2013) found that the flipped classroom allows the learning environment to be restructured in the way time, communication, and collaboration are facilitated. The classroom structure and types of feedback and collaboration were not a part of this research study but could have an impact on the effectiveness of the flipped classroom. In addition, Bull, Ferster, and Kjellstrom (2013) found that digital equity is a concern that educators must attend to during implementation of flipped classrooms. The way teachers addressed the issue of digital equity or the methods of differentiated instruction were not specifically expressed in the course of this research study. Santangelo and Tomlinson (2012) found that differentiated instruction maximizes the learning potential and outcomes for students. However, the specific ways that teachers utilized differentiated instruction were not documented. Students within college courses favor the flipped classroom because of the increased collaboration and teamwork (Clark, 2014).

Garver and Roberts (2013) highlighted the increased time for higher levels of thinking and active learning in the flipped classroom. Furthermore, Boucher, Robertson, Wainner, and Sanders (2013) noted that the flipped classroom created better results than traditional lecture as measured by student grades, surveys, and faculty feedback. Research studies in secondary education and higher education noted benefits in the implementation of the flipped classroom in the review of the literature. However, in this quantitative research study, the methods of collaboration, levels of thinking, surveys, and faculty feedback were not included in the analyses. Little quantitative research exists about the flipped classroom applied to middle school mathematics. Because of these reasons, this study can be added to the body of flipped instruction research.

In the context of this study, the flipped classroom did not have a statistically significant impact in closing the achievement gap among middle school students of various ethnic sub-populations, although the scores increased in flipped classrooms. In this study, the gap between Hispanic and White students was relatively small before the treatment was implemented. However, the gap between the African American students and their peers was wider. Closing the existing achievement gap is an important aim for educators and instructional strategies are needed for teachers to employ in the classroom that will have positive learning benefits. Another goal for this study was to investigate the impact of the flipped classroom in the instruction of pre-AP students. The results showed no significant differences among scores of pre-AP middle school mathematics students in traditional and flipped classrooms.

In this study, flipped instruction was employed for one academic year in one suburban independent school district in Texas. The first year of flipped classroom implementation showed no significant differences in achievement measured by scale scores on the state assessment (Table 4, Model 5). State assessments are summative and cover a wide range of concepts and skills. Other measures of achievement could be examined. In addition, Garver and Roberts (2013) experienced student success with the prolonged use of the flipped classroom, noting that only 5% of students scored an A on the 2006 exam in contrast to the 38% of students who scored an A on the 2011 exam. The results after five years indicate the advantages of the prolonged use of flipped instruction. Time and commitment to the change can positively impact the new implementation (Herold, Fedor, Caldwell, & Liu, 2008). It takes time for new strategies to develop and become established. Further investigation could include other measures of achievement and the prolonged use of the strategy.

Conclusion

Educational leaders are expected to be instructional leaders (Vornberg, Hickey, & Borgemenke, 2012). Choosing instructional strategies and resources to implement in the learning community is a major charge for educational leaders. The flipped classroom is an instructional approach that is gaining momentum. Additional research is required to establish the impact of prolonged used of the flipped classroom. Promising results were noted in the score gains for African American, Hispanic, and White students in the flipped classrooms (Table 4, Model 5). The scale scores of pre-AP mathematics students also noted gains in the flipped classroom (Table 4, Model 5). Although these gains were not statistically significant, this finding could have implications in our educational practices. As revealed in the review of the literature, the flipped instructional approach has many potential benefits and challenges of implementation. Further investigation is warranted in the area of the flipped classroom.

References

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine, 28*, 2037-2049.

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education, 78,* 227-236.

Bol, L., & Berry, R. Q. (2005). Secondary Mathematics Teachers' Perceptions of the Achievement Gap. *The High School Journal, 88*(4), 32-45.

Boucher, B., Robertson, E., Wainner, R., & Sanders, B. (2013). "Flipping" Texas State University's physical therapist musculoskeletal curriculum: Implementation of a hybrid learning model. *Journal of Physical Therapy Education, 27*(3), 72-77.

Brunsell, E., & Horejsi, M. (2013). Flipping your classroom in one 'take'. *The Science Teacher, 80*(3), 8.
Bull, G., Ferster, B., & Kjellstrom, W. (2013). Inventing the flipped classroom. *Learning & Leading with Technology, 40*(1), 10-11.
Chamberlin, M., & Powers, R. (2010). The promise of differentiated instruction for enhancing the mathematical understandings of college students. *Teaching Mathematics and Its Applications, 29*, 113-139.
Clark, K. R. (2014). Flipped out: A trend in radiologic science education. *Radiologic Technology, 85*(6), 685-687.
DeMaio, D. N., & Oakes, C. E. (2014). Flipping the classroom with screencasts. *Radiologic Technology, 85*(3), 340-343.
Dettmers, S., Trautwein, U., Ludtke, O., Kunter, M., & Baumert, J. (2010). Homework works if homework quality is high: Using multilevel modeling to predict the development of achievement in mathematics. *Journal of Educational Psychology, 102*(2), 467-482.
Flores, A. (2007). Examining disparities in mathematics education: achievement gap or opportunity gap? *The High School Journal, 91*(1), 29-42.
Fulton, K. P. (2012). 10 reasons to flip. *Kappan, 94*(2), 20-24.
Garver, M. S., & Roberts, B. A. (2013). Flipping & clicking your way to higher-order learning. *Marketing Education Review, 23*(1), 17-21. doi:10.2753/MER1052-8008230103
Gullen, K. & Zimmerman, H. (2013). Saving time with technology. *Educational Leadership, 70*(6), 63-66.
Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.
Herold, D. M., Fedor, D. B., Caldwell, S., & Liu, Y. (2008). The effects of transformational and change leadership on employee's commitment to a change: A multilevel study. *Journal of Applied Psychology, 93*(2), 346-357.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199-236.
Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge. Love, B., Hodge, A., Grandgenett, N., & Swift, A. (2014). Student learning and perceptions in a flipped linear algebra course. *International Journal of Mathematical Education in Science and Technology, 45*(3), 317-324.
Millard, E. (2012). 5 reasons flipped classrooms work. *University Business, 15*(11), 26-29.
Onorato, M. (2013). Transformational leadership style in the educational sector: An empirical study of corporate managers and educational leaders. *Academy of Educational Leadership Journal, 17*(1), 33-47.
Randolph, J. J., Falbe, K., Manuel, A. K., & Balloun, J. L. (2014). A step-by-step guide to propensity score matching in R. *Practical Assessment, Research & Evaluation, 19*(18), 1-6.
Rickles, J. (2011). Constructing counterfactuals in a multisite observational study using propensity score matching and multilevel modeling: An empirical example looking at the effect of 8th grade algebra across students and schools. *Society for Research of Educational Effectiveness*, 1-10.
Santangelo, T., & Tomlinson, C. A. (2012). Teacher educators' perceptions and use of differentiated instruction practices: An exploratory investigation. *Action in Teacher Education, 34*, 309-327. doi:10.1080/01626620.2012.717032
Smith, I., & Addison, C. (2013). The "new" school leader: Training instructional leaders for a new generation of teachers and learners. *Academy of Educational Leadership Journal, 17*(2), 135-140.
Stuart, E. A., & Rubin, D. B. (2008). Matching with multiple control groups with adjustment for group differences. *Journal of Educational and Behavioral Statistics, 33*(3), 279-306.
Texas Education Agency. (2010). *House Bill 3 Transition Plan*. Retrieved from
http://www.tea.texas.gov/student.assessment/hb3plan/
Texas Education Agency. (2012). *Raw Score Conversion Tables*. Retrieved from
http://www.tea.texas.gov/student.assessment/staar/convtables/
Texas Education Agency. (2013a). *2012-2013 Texas Academic Performance Reports*. Retrieved
from http://ritter.tea.state.tx.us/perfreport/tapr/2013/index.html
Thiele, H. (2013). Blended and flipped learning. *Technology & Learning, 34*(2), 44.
Thoemmes, T. (2012). *Propensity score matching in SPSS*. Retrieved from http://arxiv.org/pdf/1201.6385.pdf
Tucker, B. (2012). The flipped classroom. *Education Next, 12*(1), 82-83.
Vornberg, J. A., Hickey, W. D., & Borgemenke, A. (2012). *Texas Public School Organization and Administration: 2012*. Dubuque, IA: Kendall Hunt Publishing.