The Mathematics Enthusiast

Volume 9
Number 1 *Numbers 1 & 2*

Article 6

1-2012

Student Enrolment in Classes with Frequent Mathematical Discussion and Its Longitudinal Effect on Mathematics Achievement

Karl W. Kosko

Student Enrollment in Classes with Frequent Mathematical Discussion and Its Longitudinal Effect on Mathematics Achievement

Karl W. Kosko

University of Michigan[1]

Abstract:

Mathematical discussion has been identified as being beneficial to students' understandings of mathematics (Goos, 1995; Lee, 2006). Students in classrooms with more effective math discussion have been observed to engage more frequently in discussion (e.g. Hiebert & Wearne, 1993), but the converse is not necessarily true (e.g. Manouchehri & St. John, 2006). Utilizing hierarchical linear modeling, the present study examined student enrollment in classes with more and less frequent discussion and such enrollment's effect on mathematics achievement over time. Results indicated that students enrolled in classes that discuss math "almost every day" consistently have higher math achievement than students enrolled in classes that discuss math "never or hardly ever." These results and their implications are discussed in depth.

Keywords: Mathematical discussion, mathematics achievement, hierarchical linear modeling.

Introduction

[1] Email: kwkosko@umich.edu

The Mathematics Enthusiast, **ISSN 1551-3440, Vol. 9, nos.1&2**, pp.111-148

According to Silver, Kilpatrick, and Schlesinger (1990), "mathematics deepens and develops through communication" (p. 15). Students gain a better understanding of the meaning of mathematics when they communicate with others about it. Mathematics discussion in the classroom involves students in describing, explaining, defending, and justifying their ideas about mathematics. By doing so, mathematics discussion deepens students understanding of mathematics (Goos, 1995; Lee, 2006; Pimm, 1987) and has been shown to have a positive impact on mathematical achievement (D'Ambrosio, Johnson, & Hobbs, 1995; Grouws, 2004; Hiebert & Wearne, 1993; Koichu, Berman, and Moore, 2007; Mercer & Sams, 2006). Yet, there is evidence that discussion does not always have a positive impact on mathematics achievement (Shouse, 2001), which may imply that either discussion is not consistently effective in deepening mathematical understanding or that it is not consistently implemented to maximize its effectiveness. Perhaps this inconsistency in the research concerning the effectiveness may explain why Pimm (1987) identifies mathematical discourse as a topic that is continuously advocated by researchers but rarely implemented by teachers.

In addition to the contradictory results of research on the impact of mathematics discussion on math achievement (e.g. Mercer & Sams, 2006; Shouse, 2001), there are few examples of such research. Of the studies that focus on discussion's impact on math achievement, all are cross-sections of the samples evaluated. To date, the author has yet to find a longitudinal study to investigate the compound effects of discussion on mathematics achievement. Given that many teachers do not engage their students in mathematical discussion (Pimm, 1987), it may very well be that positive benefits of more frequent discussion may not be statistically evident within one iteration of its implementation.

The current study seeks to investigate whether students' presence in a classroom with frequent mathematical discussions has any longitudinal effect on their mathematics achievement. Students who are in classrooms where their peers frequently talk about mathematics should be more likely to be able to meaningfully and skillfully discuss mathematics than students who have not been in such classroom environments. Likewise, this ability should have a positive influence on their mathematics achievement. Therefore, students who are more exposed to classroom environments with frequent discussion about mathematics should demonstrate more growth in mathematics achievement than students who are less exposed to such classroom environments.

Mathematical Discussion and Mathematics Achievement

Describing effective teaching strategies for increasing mathematics achievement, D'Ambrosio et al. (1995) suggested that engaging students in discussions about mathematics would improve their mathematical understanding of it. One study supporting this claim was conducted by Hiebert and Weane (1993) with second grade students and teachers. Hiebert and Wearne observed six classrooms and found that two teachers observed asked students to explain and justify their mathematics significantly more than the other four teachers. In addition to being engaged in mathematical discussion more frequently, students of these two teachers had statistically significant higher gains in content knowledge than the students of the four teachers who engaged in mathematical discussion less often.

A study in Great Britain conducted by Mercer and Sams (2006) compared teachers who received training in setting up mathematical discourse environments to those who did not. Students in the treatment group engaged more frequently in self-directed discussions about mathematics, while students in the control did so less often and to a lesser degree. Similar to the findings of Hiebert and Wearne (1993), Mercer and Sams found that students in the treatment had significantly higher gains in math achievement than students in the control.

Other studies support the two previously mentioned studies' claims of mathematical discussions' positive impact on student math achievement (e.g. Koichu et al., 2007; Stigler & Hiebert, 1997). However, Shouse (2001) found that more frequent mathematical discussion had a negative impact on high school student math achievement. Shouse used a regression analysis with the 1988 National Education Longitudinal Study (NELS) dataset. The resulting coefficient was small but negative, providing a contrast to the findings of other studies. However, Kosko and Miyazaki (in press) found that the impact of the frequency that 5th graders discuss mathematics varies significantly (statistically and meaningfully) between classrooms and schools. In some schools the impact of discussion was overwhelmingly positive while in others the impact was largely negative. Additionally, the frequency students' 3rd grade classrooms engaged in math discussion increased the effect of 5th grade discussion frequency (Kosko & Miyazaki, 2009).

The results found by Kosko and Miyazaki (in press) suggest that more frequent discussion does not necessarily mean better discussion, but student exposure to more frequent discussion in the previous grade suggested that more frequent discussion may have a positive effect on math achievement over time (Kosko & Miyazaki, 2009). Hiebert and Wearne (1993) and Mercer and Sams (2006) both observed more frequent discussion on the part of students who had larger gains in math achievement. Additionally, since mathematical discussion is argued to increase math achievement (D'Ambrosio et al., 1995), it is logical to expect that more frequent student involvement in discussion is necessary for students to see the benefits of discussion.

Observed Implementation of Mathematics Discussion

There are several qualitative studies where teachers identified as implementing effective mathematical discourse are observed (e.g. Truxaw & DeFranco, 2007; Williams & Baxter, 1996; Wood, 1999). Teachers in these studies typically emphasize the characteristics of dialogic discourse when engaging students in mathematical discussion. Dialogic discourse involves both students and the teacher in developing the course of a discussion. Students are encouraged to justify and explain their reasoning while the teacher creates a positive atmosphere lacking social penalties for incorrect math answers. Additionally, students were informed why they were being asked to explain and justify mathematical ideas as well as how to go about doing it (Williams & Baxter, 1996; Wood, 1999). As encouraging as the teachers in these studies may be, it does not paint an accurate picture of how many other teachers implement mathematical discussion.

Manouchehri and St. John (2006) compared two episodes of classroom talk where there was a large degree of student participation. On the surface the two episodes appeared to be similar in that the teachers actively engaged students in the topic discussed. Yet the teachers in both classrooms acted differently in how material was explained. In one classroom the teacher explained and justified mathematical positions where in the other classroom it was the students who did so. Kazemi and Stipek (2001) found similar results in studying fourth and fifth grade classrooms. On the surface, all teachers seemed to have similar levels of discussion in their classrooms and a positive social environment for students to learn in. Results showed, however, that one set of teachers was more likely to require explanation and justification from their students than the other set of teachers. Characteristic of the teachers studied by Kazemi and Stipek (2001) was that while all four teachers asked their students to describe how they solved problems, some teachers had students discuss such descriptions while other teachers simply asked whether the class agreed or not. The lesson to be learned from these two studies is that while some teachers may seem like they are actively implementing mathematical discussion more frequently, it such discussions not contain the elements that make the discussion effective.

Contextual Effects of Mathematical Discussion

> In trying to make sense of what someone says, we never rely only on our knowledge of the basic meanings of words, or our familiarity with the grammatical constructions they use. As listeners, we always access some additional, contextual information, using any explicit guidance or hints provided by a speaker and drawing on any remembered past experience which seems relevant (Mercer, 2000, p. 44).

The above quote by Mercer (2000) demonstrates the importance of context on an individual and their discourse-related decisions. The very context a student is in not only helps define the student's interpretations of what others say but also defines the social and sociomathematical norms that the student abides by (Yackel & Cobb, 1996). The student also contributes to both social and sociomathematical norms, whether knowingly or not. This reflexive relationship is a key ingredient in what Yackel and Cobb described as the development of intellectual autonomy. "Students who are intellectually autonomous in mathematics are aware of, and draw on, their own intellectual capabilities when making mathematical decisions and judgments as they participate in these practices" (Yackel & Cobb, 1996, p. 473). Students without such autonomy rely on "pronouncements of an authority" (p. 473), such as the teacher or a textbook. Further, Yackel and Cobb emphasized the necessary involvement of students' co-development of sociomathematical norms as instrumental in their development of intellectual autonomy in the mathematics classroom. Yet it is important to note the teacher's role in guiding the development of sociomathematical norms within a classroom.

Investigating and comparing the development of sociomathematical norms in two different classrooms, Lopez and Allel (2007) noted that the way teachers go about having explanations and solution strategies validated can influence how students participate in the classroom. Lopez and Allel found that by providing students with opportunities to evaluate their peers' mathematical explanations, students became more self-regulated to engage in such actions. Similar to the findings of Lopez and Allel, McClain and Cobb (2001) found that established sociomathematical norms can provide "directionality to the students' learning…" (p. 264). Additionally, prior experiences in contexts with facilitative soicomathematical norms were found to support students' autonomous conjecture.

The appropriate development of sociomathematical norms facilitates students' development of mathematical dispositions. Yet, simply providing opportunities for students to discuss mathematics does not always yield productive or rich discussions. Sfard (2007) observed 12 and 13 year olds discussing mathematics with their teacher. During one discussion, a conflict between the teacher and students emerged but the students failed to make any attempts to rectify this conflict. Sfard characterized this breakdown of discourse as being caused by the students' lack of properly developed sociomathematical norms. Sfard provides other characterizations of "students' unawareness of what kind of argument counts as legitimate in a mathematics classroom" (p. 594).

Both Yackel and Cobb (1996) and McClain and Cobb (2001) characterized how, over time, the development of certain sociomathematical norms facilitates students' intellectual autonomy, or mathematical dispositions. Lopez and Allel (2007), along with McClain and Cobb, further characterize how exposure to such norms influences student actions. Sfard (2007) provided a description of what a lack of developed sociomathematical norms looks like in mathematical discussions. For the purposes of the current study, the literature presented here is meant to emphasize the importance of students being in such discourse environments where sociomathematical norms *can* develop.

The Current Investigation

The present study used a national dataset collected by the U.S. Department of Education. The benefits of using such a dataset include its relative size and generalizability, its taking into account of the nested nature of educational data, and the reliability of its measures. The main drawback is that the items asked of teachers, parents, administrators, and students were not items specifically tailored for a specific research interest or area. Yet, often the benefits outweigh the drawbacks and, if such drawbacks are taken into consideration, these datasets can answer research questions that could not otherwise be evaluated with smaller samplings.

Such is the case with the present study. The dataset used here was the Early Childhood Longitudinal Study (ECLS). The study collected, among other variables, the frequency teachers' classes engaged in mathematical discussion and discourse-related actions, but did not assess questions that would provide a description of these classrooms' sociomathematical norms. However, "to understand the role of any given sociomathematical norm, it [is] necessary to analyse how it [is] related to the other norms" (Lopez & Allel, 2007, p. 263). As Lopez and Allel noted, sociomathematical norms are complex and the role one sociomathematical norm plays in one classroom can be quite different in another classroom. Therefore, any quantitative assessment of sociomathematical norms would likely be an exercise in and of itself, and beyond the scope of this study. What the current study seeks to examine is not the impact of sociomathematical norms on student math achievement, but the effect of a certain, general context has on math achievement. This context can be described as classrooms where frequent discussion takes place.

Classrooms in which frequent discussion takes place may or may not have properly developed sociomathematical norms. This is evidenced by the contrasting descriptions provided by Sfard (2007) and McClain and Cobb (2001). However, student exposure to classroom environments where discussion occurs frequently may, over the course of time, allow the student to develop competencies in mathematical discussion. The development of such competencies would, undoubtedly, benefit from the teacher's purposeful guidance in co-constructing sociomathematical norms with the students. Yet, it is equally logical to conclude that given enough exposure to contexts with frequent mathematical discussion, most students will develop some level of competence in mathematical discussion. Students with higher levels of mathematical discussion ability should also have higher mathematics achievement (Mercer & Sams, 2006). This line of logic leads to the following research questions:

1. Does student presence in classrooms with frequent discussion have a general, longitudinal effect on their mathematics achievement growth?
2. Does the effect of classroom discussion frequency differ from one grade to the next?
3. Do different frequencies of classroom discussion have more positive effects on individual student math achievement than other frequencies?

Methods

Sample and Data

Data collected from the Early Childhood Longitudinal Study (ECLS) was used in this study. ECLS was designed as a longitudinal study collecting data from kindergarten students in the 1998-1999 school year through their eighth grade enrollment in 2006-2007 (NCES, 2009). In all, data was collected in kindergarten, first grade, third grade, fifth grade, and eighth grade. The current study uses data from each grade level, which included different sample sizes each year due to attrition. Items selected from teacher questionnaires of students in the sample were a primary source of data, as were student math achievement scores. Due to missing data on questionnaires and attrition, some sample sizes were reduced. The sample sizes for each year are displayed in Table 1.

Table 1

Sample Sizes for Each Year of Data.

Grade & Year	Students	Teachers
Kindergarten 1998 – 1999	11,461	1,778
1st Grade 1999 – 2000	8,939	2,276
3rd Grade 2001 – 2002	7,336	2,713
5th Grade 2003 – 2004	3,358	1,763
8th Grade 2006 – 2007	2,832	1,641

Note: Samples presented here are effective sample sizes for analysis.

As can be observed from Table 1, the students per teacher ratio decreased each year to a mere 1.73 students per teacher in grade 8. An examination of the longitudinal effect of frequent classroom discussion was therefore conducted at the individual level. This decision and its implications are discussed in the following section.

Measures

Dependent variable.

The dependent variable, or outcome variable, in the current study is student mathematics achievement as measured with a standardized cognitive domain test (NCES, 2009). Versions of this assessment were administered in each year of data collection of ECLS and included a variety of math content. The mathematics cognitive domain test scores were standardized using Item Response Theory scale scores (IRT scores). IRT scores utilize student item response patterns to obtain a scale score that represented their content knowledge and, therefore, their ability. One of the advantages of IRT scale scores is their comparability to other achievement measures observed at different time points. IRT scores measured in the spring of each year in data collection were used as the outcome measure of the longitudinal analysis (*Math_IRT*) and the cross section analyses (*GK_IRT, G1_IRT, G3_IRT, G5_IRT, G8_IRT*), representing grades K through 8, respectively. Descriptive statistics of these variables can be found in Table 2.

Table 2

Descriptive Statistics for Students Math Achievement

	Range	Mean	S.D.	*N*
GK_IRT	11.57 – 112.51	35.98	11.83	11,461
G1_IRT	13.44 – 132.49	60.56	18.32	8,939
G3_IRT	37.47 – 166.25	97.12	25.14	7,336
G5_IRT	50.86 – 170.66	119.87	26.11	3,358
G8_IRT	66.26 – 172.20	138.05	23.75	1,641

Note: Statistics were weighted using appropriate cross-section weights

Independent variable.

Each spring, the teachers of students enrolled in the ECLS study completed a questionnaire asking questions regarding teacher background, instructional practices, observations of the child participant, and observations of the child participant's class. Of interest in the current investigation were items regarding mathematical discussion. The question asked for each grade level assessed is shown in Table 3.

Table 3.

ECLS Items Assessing Discussion Frequency.

Grade	Question	Responses
K, 1	How often do children in this class do each of the following MATH activities? -Explain how a math problem is solved. (NCES, 1999, p. 15; NCES, 2000, p. 20)	Never; Once a Month; Two or Three Times a Month; Once or Twice a Week; Three or Four Times a Week; Daily
3	How often do children in your class engage in the following?	Never or Hardly Ever;

	-Discuss solutions to mathematics problems with other children (NCES, 2002, p. 20)	Once or Twice a Month; Once or Twice a Week; Almost Everyday
5	How often does the child identified on the cover of this questionnaire engage in the following as part of mathematics instruction? -Discuss solutions to mathematics problems with other children. (NCES, 2004, p. 6)	Never or Hardly Ever; Once or Twice a Month; Once or Twice a Week; Almost Everyday
8	How often do the students in this class engage in the following? -Discuss their solutions to mathematics problems. (NCES, 2007, p. 11)	Never or Hardly Ever; Once or Twice a Month; Once or Twice a Week; Almost Everyday

As shown in Table 3, responses to discussion items in third, fifth, and eighth grades were on a 4 point scale. Therefore the Kindergarten and first grade responses were recoded to match the outcomes of later grades. “Never” was recoded to match “never or hardly ever;” “once a month” and “two or three times a month” were recoded to match “once or twice a month;” “Once or twice a week” was matched to “once or twice a week;” and “three or four times a week” and “daily” were matched to “almost everyday.”

The fifth grade item was assessed for the participating student, whereas all other discussion related items were assessed of the participating student’s class. Thus, this item was recoded to reflect classroom frequency rather than individual frequency. An aggregate variable was created for each classroom and then these means were rounded to the nearest whole number to match the 4 point scale of the other items.

After the recoding of discussion items outlined in the previous paragraphs was conducted, it was decided that frequencies of discussion would be dummy coded with *Never or Hardly Ever* as the reference group. *Almost Everyday* became *disc_daily* (1 = Almost Everyday, 0 = all other frequencies); *Once or Twice a Week* became *disc_weekly* (1 = Once or Twice a Week, 0 = all other frequencies); and *Once or Twice a Month* became *disc_monthly* (1 = Once or Twice a Month, 0 = all other frequencies). These new variables were formatted as within-student variables for longitudinal analysis and student-level variables for cross sectional analyses. Descriptive statistics of the discussion variables are displayed in Table 4.

The variables *disc_daily*, *disc_weekly*, and *disc_monthly* were defined so that they represented student enrollment in classes with more or less frequent discussion. This is an important distinction to make. The variables, as defined in this study, do not represent student frequency of discussion or classroom frequency of discussion. Since it was defined as a within-persons variable and student-level variable (for longitudinal and cross section analysis respectively), the dummy coded variables are characterized as enrollment. A similar type of variable assignment can be likened to student enrollment in a specific content level course. Such was done by Ma and Wilkins (2008) who used math course type as a student level variable which represented student enrollment in the course rather than the items' original description of course type.

Table 4

Descriptive Statistics for Recoded Discussion Variable by Grade.

Grade	Level of Discussion	Enrollment Frequency	Weighted Statistics*
K	Never or Hardly Ever	1052 (9.2%)	$\bar{X}$ = 1.93
	Once or Twice a Month	3047 (26.6%)	S.D. = .99
	Once or Twice a Week	3204 (28.0%)	*n* = 11,461
	Almost Everyday	4158 (36.3%)	
1	Never or Hardly Ever	65 (0.7%)	$\bar{X}$ = 2.55
	Once or Twice a Month	913 (10.2%)	S.D. = .70
	Once or Twice a Week	2244 (25.1%)	*n* = 8,939
	Almost Everyday	5717 (64.0%)	
3	Never or Hardly Ever	518 (7.1%)	$\bar{X}$ = 2.03
	Once or Twice a Month	1510 (20.6%)	S.D. = .90
	Once or Twice a Week	2835 (38.6%)	*n* = 7,336
	Almost Everyday	2475 (33.7%)	
5	Never or Hardly Ever	233 (6.9%)	$\bar{X}$ = 1.97
	Once or Twice a Month	540 (16.1%)	S.D. = .99
	Once or Twice a Week	1338 (39.8%)	*n* = 3,358
	Almost Everyday	1247 (37.1%)	
8	Never or Hardly Ever	60 (2.1%)	$\bar{X}$ = 2.61
	Once or Twice a Month	140 (4.9%)	S.D. = .66
	Once or Twice a Week	757 (26.7%)	*n* = 2,832
	Almost Everyday	1875 (66.2%)	

**Cross-Section Weights C2CW0, C4CW0, C5CW0, C6CW0, & C7CW0 were used for each respective grade level. Means are based off of the following coding scheme (0 = Never or Hardly Ever; 1 = Once or Twice a Month; 2 = Once or Twice a Week; 3 = Almost Everyday).*

Covariates.

Covariates included at the individual level for both longitudinal and cross section analyses included student gender (*dFemale*) and race/ethnicity (*dBlack*, *dHispanic*, *dAsian*, *dOther*). Socio-economic status (*SES*) was included at the within-student level for the longitudinal analysis since it can vary from year to year, but was included at the student level for the cross section analyses. *SES* was calculated at the household level and included the following components: father/male guardian's education; mother/female guardian's education; father/male guardian's occupation; mother/female guardian's occupation; and household income. For further details on how SES was computed, see NCES, 2009, p. 7-23.

Gender and race/ethnicity variables were included at the student level for both the longitudinal and cross section analyses. Gender (dFemale) was dummy coded to compare to male studentds (0 = male, 1 = female). Each race/ethnicity variable (*dBlack*, *dHispanic*, *dAsian*, *dOther*) were similarly dummy-coded as to compare to white students. For example, *dBlack* was coded such that 1 = Black, and 0 = non-Black. Descriptive statistics for all covariates for each grade level are presented in Table 5.

Table 5.

Descriptive Statistics of Covariates.

	Kindergarten	1st Grade	3rd Grade	5th Grade	8th Grade
dFemale	$\bar{X}$ = .49 SD = .50	$\bar{X}$ = .49 SD = .50	$\bar{X}$ = .49 SD = .50	$\bar{X}$ = .49 SD = .50	$\bar{X}$ = .49 SD = .50
dBlack	$\bar{X}$ = .18 SD = .38	$\bar{X}$ = .16 SD = .37	$\bar{X}$ = .15 SD = .35	$\bar{X}$ = .14 SD = .35	$\bar{X}$ = .12 SD = .32
dHispanic	$\bar{X}$ = .10 SD = .30	$\bar{X}$ = .09 SD = .29	$\bar{X}$ = .09 SD = .28	$\bar{X}$ = .10 SD = .30	$\bar{X}$ = .08 SD = .28
dAsian	$\bar{X}$ = .05 SD = .21	$\bar{X}$ = .04 SD = .21	$\bar{X}$ = .04 SD = .20	$\bar{X}$ = .05 SD = .21	$\bar{X}$ = .04 SD = .20
dOther	$\bar{X}$ = .05 SD = .21	$\bar{X}$ = .04 SD = .20	$\bar{X}$ = .04 SD = .20	$\bar{X}$ = .05 SD = .21	$\bar{X}$ = .05 SD = .20
SES	$\bar{X}$ = .04 SD = .78	$\bar{X}$ = .04 SD = .79	$\bar{X}$ = .03 SD = .78	$\bar{X}$ = .03 SD = .78	$\bar{X}$ = .04 SD = .78

Analysis

A two level hierarchical linear model (HLM-2) was used in the current study. HLM can be conceptually described as a "hierarchical system of regression equations" (Hox, 2002, p. 11).

HLM was employed both for the cross section analyses of each grade level and for the longitudinal analysis across grade levels. For the cross section models, I considered students as nested within classrooms (or teachers). HLM-2 allows us to explain this nested relationship by having separate regression equations for each classroom and an additional regression equation that examines the classroom-level data. For the longitudinal model, we considered variables measured at differing grade levels as being nested within the individual. Certain variables (e.g. *discuss*, *SES, Math_IRT*) change for the individual over time and are therefore nested aspects of the individual. For longitudinal models, HLM-2 allows us to examine the slope of growth as attributed to time and other factors. While the regression equations in the cross section models allow for interpretations of effect and/or impact on math achievement, the regression equations in the longitudinal HLM-2 model allows for the interpretation of coefficients as general changes over time in the effect/impact itself.

Specification of the Cross-Section Models

Five separate HLM-2 cross-section models were run using HLM6 software (Raudenbush, Bryk, & Congdon, 2007). Students were considered nested within classrooms (or teachers). Therefore, level-1 represented student level variables and level-2 represented classroom level variables. For the purpose of comparison, the model specifications were the same for each model:

$$\left(Math_{IRT_{K,1,3,5,8}}\right)_{ij} = \beta_{0j} + \beta_{1j}\left(disc_{daily_{K,1,3,5,8}}\right)_{ij} + \beta_{2j}\left(disc_{weekly_{K,1,3,5,8}}\right)_{ij} + \beta_{3j}\left(disc\_monthly_{K,1,3,5,8}\right)_{ij} + \beta_{4j}\left(Prior\_Math\_IRT\right)_{ij} + \beta_{5j}\left(dFemale\right)_{ij} + \beta_{6j}\left(SES_{K,1,3,5,8}\right)_{ij} + \beta_{7j}\left(dBlack\right)_{ij} + \beta_{8j}\left(dHispanic\right)_{ij} + \beta_{9j}\left(dAsian\right)_{ij}\beta_{10j}\left(dOther\right)_{ij} + r_{ij}$$

In the level-1 model displayed above, $(Math\_IRT_{K,1,3,5,8})_{ij}$ represents the Math IRT score student *i* achieved in classroom *j* each spring given the specific grade level (i.e. K, 1, 3, 5, 8). β_{0j} represents the average grade-specific math IRT score for white male students enrolled in classrooms that discuss mathematics *Never or Hardly Ever*, adjusted for prior achievement and *SES*. β_{1j}, β_{2j}, and β_{3j} represent the effect of student enrolment type on their mathematics IRT score for that particular grade. β_{4j} represents the association of prior achievement with students' spring math IRT scores. Spring math IRT scores from the previous measure in the data were used as the prior achievement measure for grades 1, 3, 5, and 8. For Kindergarten, a math IRT score obtained in the Fall of 1998 was used as the measure for prior achievement. β_{5j} is the gender effect and β_{6j} represents the effect of SES. Finally, the coefficients β_{7j}, β_{8j}, β_{9j}, and β_{10j} represent the effects of race/ethnicity.

The main reason for our use of cross-section analyses in each grade level was to have an additional perspective on the effect of students being enrolled in courses with frequent discussion from grade to grade. Therefore, classroom level variables were not examined. However, the slopes of β_{1j}, β_{2j}, and β_{3j} were set to vary randomly at level-2. This allowed for differences in the effect of student enrollment to vary between classrooms.

Specification of the Longitudinal Model

HLM-2 was used for the longitudinal analysis. Students were the level-2 grouping factor and level-1 was specified as within-student measures (i.e. *disc_daily, disc_weekly, disc_monthly, SES, Math_IRT*). Additional interaction effects between grade level and the discussion variables were also included. Therefore, level-1 was a set of separate regression equations, one for each student (Hox, 2002). Students' longitudinal math IRT scores were regressed, at level-1, onto their grade level, class discussion enrollment, and their socio-economic status.

<u>Level-1</u>:

$$
\begin{aligned}
Math_{IRT\,ti} = {} & \pi_{0i} + \pi_{1i}(grade)_{ti} + \pi_{2i}(disc_{daily})_{ti} + \pi_{3i}(disc_{weekly})_{ti} \\
& + \pi_{4i}(disc_{monthly})_{ti} + \pi_{4i}(grade \times disc_{daily})_{ti} \\
& + \pi_{4i}(grade \times disc_{eekly})_{ti} + \pi_{4i}(grade \times disc_{monthly})_{ti} \\
& + \pi_{3i}(SES)_{ti} + R_{ti}
\end{aligned}
$$

<u>Level-2</u>:

$$
\begin{aligned}
\pi_{0i} &= \beta_{00} + \beta_{01}(dFemale)_i + \beta_{02}(dBlack)_i + \beta_{03}(dHispanic)_i + \beta_{04}(dAsian)_i \\
&\quad + \beta_{05}(dOther)_i + \mu_{0i} \\
\pi_{1i} &= \beta_{10} + \beta_{11}(dFemale)_i + \beta_{12}(dBlack)_i + \beta_{13}(dHispanic)_i + \beta_{14}(dAsian)_i \\
&\quad + \beta_{15}(dOther)_i + \mu_{1i} \\
\pi_{2i} &= \beta_{20} \\
\pi_{3i} &= \beta_{30} \\
\pi_{4i} &= \beta_{40} \\
\pi_{5i} &= \beta_{50} + \beta_{51}(dFemale)_i + \beta_{52}(dBlack)_i + \beta_{53}(dHispanic)_i + \beta_{54}(dAsian)_i \\
&\quad + \beta_{55}(dOther)_i + \mu_{5i} \\
\pi_{6i} &= \beta_{60} + \beta_{61}(dFemale)_i + \beta_{62}(dBlack)_i + \beta_{63}(dHispanic)_i + \beta_{64}(dAsian)_i \\
&\quad + \beta_{65}(dOther)_i + \mu_{6i} \\
\pi_{7i} &= \beta_{70} + \beta_{71}(dFemale)_i + \beta_{72}(dBlack)_i + \beta_{73}(dHispanic)_i + \beta_{74}(dAsian)_i \\
&\quad + \beta_{75}(dOther)_i + \mu_{7i} \\
\pi_{8i} &= \beta_{80}
\end{aligned}
$$

π_{0i} represents the math IRT score of student *i* at the initial measurement, spring 1999. π_{1i} represents the effect of grade level on students' math IRT score. In other words, π_{1i} can be viewed as the natural effect of time on increasing math achievement. π_{2i}, π_{3i}, and π_{4i} are the general effects of student enrollment in classrooms with different levels of frequent math discussion. However, each of these variables disregards the effect of time and individual. Therefore, while these variables are statistically necessary for a rigorous analysis, they are not meaningfully useful for the research questions posed in the current study. Therefore, the interaction effects π_{5i}, π_{6i}, and π_{7i} where included. Each of these represents the general change enrollment in classes with more or less frequent math discussion had on student growth in math achievement as they progressed in grade level. π_{3i} represents the effect of SES on math achievement over time.

The slopes of *grade*, $grade \times disc\_daily$, $grade \times disc\_weekly$, and $grade \times disc\_monthly$ were set to vary randomly at level-2. Setting *grade* to vary randomly between individuals is a logical decision since different individuals will likely experience different rates of growth in their mathematics achievement over their schooling. Setting $grade \times disc\_daily$, $grade \times disc\_weekly$, and $grade \times disc\_monthly$ to vary randomly was done to see if the interaction effect between enrollment and grade level varied between individuals. In the model equation displayed above, only the level-1 variables that were set to vary randomly were modeled at level-2. Additionally, each of these slopes were regressed on gender and race/ethnicity to examine whether there were significant effects for these factors.

Results

Cross-Section Results

While it is customary to present baseline results for all HLM models constructed, for the sake of brevity, only the final models for the cross-section analyses are presented here. Since the main purpose of this study was to examine the longitudinal effect of enrollment in classrooms with different frequencies of discussion, the cross-section results provide additional information to the longitudinal results, while not being a major point of focus. Results from cross-section analysis of each grade level yielded varying results. These results are presented in Table 6, with the shaded rows representing the effects of the independent variables of interest. In general, the intercept for each grade level was statistically significant from zero and was larger than the preceding grade level. Enrollment in classes with daily math discussion was found to be positive in every grade level and statistically significant from *Never or Hardly Ever* in each grade but 3rd grade. Enrollment in classes with weekly discussion was statistically significant only in grades K and 1 but was near significant in grade 5 (p = .085). Enrollment in classes with monthly discussion was only statistically significant in kindergarten. An overall look at the impact of enrollment by grade level illustrates that in grades K and 1 a generally positive relationship between the frequency a class discusses mathematics and the enrolled student's math achievement was found. However, this trend becomes convoluted by grade 3. In 3rd grade, all enrollment types are statistically similar, but weekly discussion had a larger effect than daily discussion. In 5th grade, monthly discussion had a larger effect than weekly and in 8th grade monthly discussion had a larger effect than weekly and daily discussion. Yet, enrollment in classes with monthly discussion was not statistically significant in grades 5 or 8. This appears to be due to large standard error for 5th grade (S.E. = 2.78) and 8th grade (S.E. = 2.65) at level-1. A similar finding appears for enrollment in classes with both daily and weekly discussion. In 3rd grade, the standard error for each enrollment type was larger than the level-1 coefficients. Typically, such large level-1 standard errors would indicate a possibly significant amount of variability at level-2 as well. Due to loss in degrees of freedom, setting the dummy coded variables' slopes as random at level-2 would not allow the model to converge. Therefore, such a possibility was not able to be examined in detail. What can be concluded from the cross section analysis is that student enrollment in classes with daily discussions about mathematics will generally predict higher levels of math achievement. This is true in grades K, 1, 5, and 8. Enrollment in classes with other levels of discussion can predict higher math achievement than enrollment in classes with little or no discussion, but such effects appear to vary since the standard errors for these effects were often quite large in later grades.

Table 6.

Results of Final Models by Grade for Cross-Section Analyses.

Estimated Fixed Effects

	K	1	3	5	8
β_{0j}, intercept	36.22**	59.36**	101.70**	120.09**	137.48**
β_{1j}, disc_daily	1.81**	5.53**	1.09	5.54*	3.89*
β_{2j}, disc_weekly	1.71**	4.51*	1.11	4.48	3.47
β_{3j}, disc_monthly	1.08**	3.40	0.13	4.55	4.42
β_{4j}, Prior IRT	1.04**	1.07**	0.97**	0.82**	0.75**
β_{5j}, dFemale	-0.34*	-0.82**	-3.61**	-1.16*	1.50
β_{6j}, SES	0.83**	2.06**	4.26**	2.54**	1.78**
β_{7j}, dBlack	-1.59**	-2.86**	-4.58**	-3.98**	-0.61
β_{8j}, dHisp	-0.87**	-0.34	-0.22	1.43	-1.69
β_{9j}, dAsian	0.62	0.46	0.89	2.76	1.90
β_{10j}, dOther	-0.65	-2.76**	-1.49	-1.56	1.39

Estimation of Variance Components

	K	1	3	5	8
μ_{00}	2.16**	3.71**	6.40**	6.27**	6.11**
r_{ij}	6.12	10.34	12.79	10.39	9.09

**p<.05, **p<.01*

Prior achievement and SES were found to be statistically significant for each grade level. The effect of prior achievement decreased for each grade level, while the effect of SES increased to a maximum at grade 3 but then decreased substantially after that point. Students who were black had a statistically significant negative effect in each grade except 8th grade. Students of other ethnicities had a statistically significant negative effect in grade 1, but were otherwise statistically similar to Caucasian students. The effect of female gender was statistically significant in every grade level and was negative through 5th grade. However, in 8th grade, girls appeared to outperform boys in general. While the findings of the covariates are interesting in and of themselves, it is the effect of *disc_daily*, *disc_weekly*, and *disc_monthly* that are of primary interest in the current investigation.

Longitudinal Results

Baseline model results.

Results of the baseline model are displayed in Table 7. Students were found to vary significantly on their initial math achievement ($var(\rho_{0j})$ = 165.78, *p* < .01), but students were not found to vary significantly in their rate of growth from grade to grade ($var(\rho_{1j})$ = 1.24, *p* > .50). However, the growth rate was found to be statistically significant from zero (π_{10} = 13.14, *p* < .01), meaning that on average, students' math IRT scores increased 13.14 points every grade level.

Table 7

Results of Baseline Model

Estimation of Fixed Effects

Fixed Effect	Coefficient	Std. Error	t-ratio	df	p-value
Intercept, π_{00}	48.46	.19	260.54	7305	.000
grade, π_{10}	13.14	.03	453.71	7305	.000

Estimation of Variance Components

Random Effect	SD	Variance Component	df	Chi-Square	p-value
Intercept, ρ_{0j}	12.88	165.78	7327	15654.76	.000
grade, ρ_{1j}	1.24	1.55	7327	7250.00	>.50
Level-1 effect, r_{tj}	15.23	231.82	-	-	-

Final model results.

The results of the final model are presented in Table 8 and Table 9. Similar to the baseline model results, the intercept was found to vary significantly between students ($var(\rho_{0i})$ = 82.85, *p* < .01), but now the effect of grade level was found to vary significantly as well ($var(\rho_{1i})$ = 4.47, *p* < .01). Similar to the cross section analyses, the HLM model would not converge with any of the dummy coded interactions (*grade*, $grade \times disc\_daily$, $grade \times disc\_weekly$, $grade \times disc\_monthly$) set as random at level-2. Therefore, these coefficients were fixed at level-2.

Table 8.

Level-1 Results of Longitudinal Final Model

Estimation of Fixed Effects.

Fixed Effect	Coefficient	S.E.
π_{00}, intercept	41.71**	.60
π_{01}, dFemale	-1.41**	.34
π_{02}, dBlack	-8.52**	.52
π_{03}, dHispanic	-6.97**	.59
π_{04}, dAsian	-1.20	.86
π_{05}, dOther	-6.40**	.83
π_{10}, grade	17.48**	.25
π_{11}, dFemale	-.26	.30
π_{12}, dBlack	-1.03*	.46
π_{13}, dHispanic	-.76	.50
π_{14}, dAsian	1.89*	.74
π_{15}, dOther	-2.11**	.68
π_{20}, disc_daily	9.18**	.59
π_{30}, disc_weekly	7.49**	.61
π_{40}, disc_monthly	1.29*	.63
π_{50}, grade*daily	-3.72**	.26
π_{51}, dFemale	-.09	.30
π_{52}, dBlack	-.48	.46
π_{53}, dHispanic	1.05*	.51
π_{54}, dAsian	-.90	.74
π_{55}, dOther	1.43*	.69
π_{60}, grade*weekly	-1.85**	.27
π_{61}, dFemale	-.26	.31

π_{62}, dBlack	-1.00*	.47
π_{63}, dHispanic	.80	.52
π_{64}, dAsian	-1.38	.76
π_{65}, dOther	.28	.71
π_{70}, grade*monthly	.40	.39
π_{71}, dFemale	-.04	.42
π_{72}, dBlack	-1.82	.55
π_{73}, dHispanic	1.15	.85
π_{74}, dAsian	-1.76	1.14
π_{75}, dOther	-.44	.72
π_{30}, SES	5.62**	.21

Table 9

Estimation of Variance Components for Longitudinal Final Model.

Random Effect	Variance	df
ρ_{00}, intercept	82.85**	6785
ρ_{10}, grade	4.47**	6785
r_{ij}, level-1 error	190.35	-

Discussion variables.

The variables *disc_daily*, *disc_weekly*, and *disc_monthly* were each found to be statistically significant (π_{20} = 9.18, p < .01; π_{30} = 7.49, p < .01; π_{40} = 1.29, p < .05), indicating that, disregarding time, students enrolled in classes with more frequent discussion had higher math IRT scores than students enrolled in classes that never have discussions about mathematics. While it is tempting to regard these specific results as promoting the use of math discussion, these results should not be interpreted as such. Since *disc_daily*, *disc_weekly*, and *disc_monthly* do not acknowledge differences in math achievement due to grade level, some of the differences in the effect of enrollment in this regard may be due to certain students being enrolled in such classes in, say, 8th grade while others may be enrolled in classes with little or no discussion in, say, kindergarten or 1st grade. While the distribution of frequencies displayed in Table 4 suggest the reverse may be more typical, we cannot make the assumption that this is the case. Neither can we make the assumption that the differences found for *disc_daily*, *disc_weekly*, and *disc_monthly* are due simply to differences in grade level. The variables themselves, quite simply, cannot be interpreted in a way meaningful to the questions addressed in this study. Their purpose in the model are to moderate effects of the variables that can be more meaningfully interpreted.

Interaction variables for discussion and grade level.

The interaction effects were found to have intriguing results. $grade \times disc\_daily$ was found to be statistically significant (π_{60} = -3.72, p < .01), as was $grade \times disc\_weekly$ (π_{50} = -1.85, p < .01). $grade \times disc\_monthly$ was not found to be statistically significant (π_{70} = .40, p = .30). These results should be interpreted with care. The negative coefficient found for $grade \times disc\_daily$ indicates that, in comparison to enrollment in classes with little or no math discussion, the effect of enrollment in classes with daily math discussions decreases as students progress in school. Therefore, enrollment in classes with daily discussion is more associated with student math achievement in earlier grades than it is in later grades. A similar interpretation can be made for $grade \times disc\_weekly$. The impact of enrollment in classes with monthly discussion over time does not appear to be statistically different from that of enrollment in classes with little or no discussion. Again, the coefficients found for $grade \times disc\_daily$, $grade \times disc\_weekly$, and $grade \times disc\_monthly$ describe how the impact of enrollment changes over time, not the impact itself. Results from the cross section analyses indicate that enrollment in classes with daily discussion typically predicts higher math achievement scores than enrollment in classes with little or no math discussion. The results described in this longitudinal analysis suggest that while enrollment in classes with daily math discussion is more effective in increasing students' math achievement, this effectiveness decreases as students progress through school.

Level-1 and level-2 covariates.

SES, which was added as a covariate at level-1, was found to be statistically significant (π_{30} = 5.62, $p < .01$), indicating that students with higher SES improved their math IRT scores at a higher rate than other students. Examination of student level variables show that females tended to have lower initial math IRT scores than boys (π_{01} = -1.41, $p < .01$). Blacks, Hispanics, and students of other ethnicities tended to have lower initial math IRT scores than white students (π_{02} = -8.52, $p < .01$; π_{03} = -6.97, $p < .01$; π_{05} = -6.40, $p < .01$). Black students and students of other ethnicities tended to have slower natural growth in achievement than white students (π_{12} = -1.03, $p < .05$; π_{15} = -2.11, $p < .01$), while Asian students tended to have higher natural growth than white students (π_{14} = 1.89, $p < .05$). Hispanic students and students of other ethnicities tended to see less of a decrease in $grade \times disc\_daily$ than white students ($\pi_{33} = 1.05, p < .05$; π_{35} = 1.43, $p < .05$), but this relationship did not hold for enrollment in classes with other frequencies of math discussion. Black students tended to see more of a decrease in $grade \times disc\_weekly$ and $grade \times disc\_monthly$ than white students (π_{62} = -1.00, $p < .05$; π_{72} = -1.82, $p < .05$). While these results are intriguing, it is difficult to say what may or may not cause the differences found here. It may be that black students experience some form of social inequality when enrolled in classes with less frequent discussion, but it is strange that a similar relationship would not be found of other minority students. The results for Hispanic students and students of other ethnicities is equally perplexing.

Overview of Cross Section and Longitudinal Results

Figure 1 illustrates the growth, by grade, of student math achievement by enrollment in classes with more or less frequent discussion. As one can observe, classes that never or hardly ever discuss mathematics consistently perform less well than classes that have discussion of any frequency. The differences between grade level appear to be between the actual frequency of engaging or not engaging in discussion. This graphical analysis could provide clues for the different results found in both the cross-sectional and longitudinal analyses.

Math IRT by Grade and Class Discussion Frequency

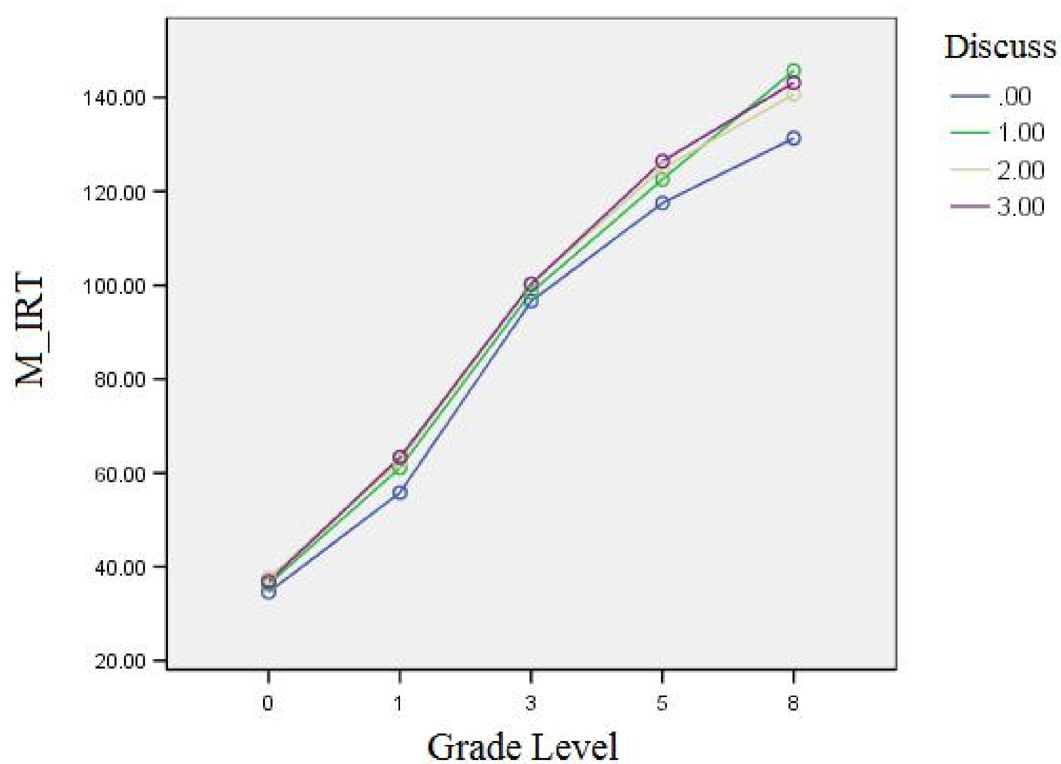

Figure 1. Graph of Math Achievement and Class Discussion Frequency by Grade Level.

The cross section analyses illustrated that the standard error found in each grade level had an effect on whether certain means were found to be statistically significant from others. For example, even though students who were enrolled in 8th grade classes with monthly discussions tended to have higher math IRT scores than students in classes with daily math discussions (see Figure 1), the degree of variance for *disc_monthly* characterized it as statistically similar to *disc_never*. This suggests that *disc_daily* is a more consistent predictor of higher math IRT scores than *disc_monthly* in 8th grade, even though *disc_monthly* had a higher mean. Similar results were found in 5th grade and 3rd grade. These results suggest that while enrollment in classes that discuss math more frequently is often more beneficial than enrollment in classes with no discussion, there is a large amount of variability in the impact of such enrollment.

The longitudinal analysis found that, over the course of time in school, the impact of *disc_daily* and *disc_weekly* affects math achievement growth at a slower rate than enrollment in classes that never discuss mathematics. When examining these results in combination with those of the cross sectional analyses, an intriguing picture begins to form. Students' enrollment in classes with more frequent discussion generally has a more positive impact on their math achievement than enrollment in classes with little to no discussion. This was found to be true in each grade level examined except 3rd grade, and these results took into account prior achievement, race/ethnicity, gender, and SES. Yet, even though enrollment in classes with daily discussion was found to consistently be a better predictor of higher math achievement than enrollment in classes with no discussion, the effect associated with *disc_daily* tended to slow down as students progressed in schooling. In other words, enrollment in classes with more frequent math discussion was more beneficial at earlier grades than later grades. Therefore, enrollment in classes with daily math discussions was found to be generally more beneficial to student achievement than enrollment in classes with little or no discussion about math, but such enrollment is relatively more effective in earlier grades than in later grades.

Discussion

The results of the present study are significant in two distinctive ways. First, cross sectional analyses found that, in general, student enrollment in classes with daily discussions about mathematics consistently outperformed students in classes with little or no discussion. A high standard error in 3rd grade prevented a statistically significant result in this regard, but in grades K, 1, 5, and 8 this trend held true. Second, the effectiveness of enrollment in classes with frequent discussion decreased as students progressed to later grade levels. Therefore, while student math achievement scores tended to benefit from enrollment in classes with daily or weekly math discussions in any grade, the size of this effect decreased when compared to the effect of being enrolled in classes with little to no discussion. While these findings are highly significant in and of themselves, it is important to remember that it was student enrollment in classes with frequent math discussion that was evaluated; not student discussion itself. Therefore, it was the effect of a context for mathematical discussions that was examined in the current study. This distinction should be considered at the forefront of the discussion that follows.

The classroom context a student is present in is an important contributor to a student's discourse-related decisions (Mercer, 2000). The context examined in the current investigation was classrooms with different frequencies of mathematical discussion. Classrooms with effective mathematical discussion practices have been observed to have more frequent discussions about mathematics (e.g. Truxaw & DeFranco, 2007; Williams & Baxter, 1996; Wood, 1999). However, the converse is not necessarily true (Kazemi & Stipek, 2001; Manouchehri & St. John, 2006). The present study took this into consideration, but given the nature of the data used, I used the frequency of discussion as a measurement rather than quality of discussion. This is not a weakness in the study's design, but rather a distinctive perspective of a certain population. Despite several qualitative studies emphasizing that mere frequency of discussion is not an indicator of effective discussion, many teachers undoubtedly use this approach. The findings of Kosko and Miyazaki (2009; 2011) suggest this may be the case. Kosko and Miyazaki (2009) observed that a significant amount of variance between classrooms and schools in the impact of more frequent student discussion on 5th grade math achievement could be explained by enrollment in 3rd grade classes with higher frequencies of math discussion. The results suggested that discussion was more effective in some classrooms and schools than others. Yet, when one examines Table 3 in the present study, we can see that in each grade well over half of classrooms have discussions about math more than once a week. This prevalence in frequency of discussion, which was evaluated in the current study, suggests that a large number of students are in classrooms with less effective discussion practices.

The results presented here suggest that even with a large amount of variability, student exposure to contexts with daily math discussions has a large and positive impact on their math achievement. Described another way, whether discussion practices are likely to be more or less effective, in general, a student enrolled in a class with daily math discussions will have larger gains in math achievement than a similar student enrolled in a class with little or no discussions about mathematics. This relationship was found to be true and statistically significant in grades K, 1, 5, and 8, and is considered to be highly significant. The potential implications of this finding suggests that while more frequent discussion does not equate to better quality discussion (Kazemi & Stipek, 2001), the context of classrooms with frequent math discussions indirectly or directly improve students' math achievement.

When one conjectures about what types of social or sociomathematical norms may exist in the classrooms with more frequent discussion, we might consider that such classrooms would be likely to have more caring and supportive environments and would involve students in the co-creation of sociomathematical norms. However, we know this is not always the case (e.g. Kazemi & Stipek, 2001; Sfard, 2007). In fact, the large amount of variance in the impact of frequent discussion found in certain grade levels in the current study suggests that this is not the case in a large number of classrooms and schools. Additional findings from other studies on discussion and math achievement suggest that the way students engage in mathematical discussion is important (Hiebert & Wearne, 1993; Mercer & Sams, 2006). However, both these studies also suggested that students who engaged in discussion about math more frequently also showed higher gains in math achievement. The current study supports these latter findings.

The findings of the current study related to the decreased impact of *disc_daily* and *disc_weekly* were surprising. Additionally, the author is at a loss for a possible reason to explain this decreasing impact. There are a number of possibilities that might be explored. First, do teachers in earlier grades facilitate mathematical discussions differently than teachers of later grades? If so, is one method better than another? It is quite possible that teachers of earlier grade levels scaffold student engagement in mathematical discussions better than teachers of later grade levels, thereby accounting for the decrease in impact found in the current study. Another possibility is that younger students may simply be more receptive to learning how to discuss mathematics than older students who may have internalized a more traditional view of mathematics and the mathematics classroom. No matter what the cause in the decrease in impact is, the results of the current study suggest two implications in this regard. First, this decrease in impact should be further studied to see what possible causes it has. Second, mathematical discussion should be encouraged early and often, as it is relatively more effective in earlier grades and generally effective in improving math achievement throughout schooling.

D'Ambrosio et al. (1995) outlined mathematical discussion as a means of increasing math achievement. Certain studies uphold this claim (Hiebert & Wearne, 1993; Mercer & Sams, 2006). The results presented here suggest that mathematical discussion does have a positive effect on students' mathematics achievement, but this effect is higher in earlier grades than in later grades. Additionally, there is a large degree of variability in the effect of enrollment in classes with more or less frequent discussion, which could infer that discussion in some classrooms has less of an impact on students' math achievement than discussion in other classrooms. Therefore, while more frequent mathematical discussion in the math classroom appears to have a generally beneficial effect on mathematics achievement, the practical implications of this study suggest that any incorporation of mathematical discussion should be accompanied with appropriate mathematical discourse practices.

References

D'Ambrosio, B., Johnson, H., & Hobbs, L. (1995). Strategies for increasing achievement in mathematics. In R.W. Cole (Ed.), *Educating everybody's children: Diverse teaching strategies for diverse learners* (121 – 138). Alexandria, VA: Association for Supervision and Curriculum Development.

Goos, M. (1995).*How do you know when you understand? Using explanation to monitor and construct mathematical understanding*. Paper presented at the Postgraduate Research Conference Commemorating 50 Years of Educational Studies at the University of Queensland, Brisbane, Queensland, Australia.

Grouws, D.A. (2004). Mathematics. In G. Cawelti (Ed.), *Handbook of research on improving student achievement* (3rd Ed.) (pp. 160 – 178). Arlington, VA: Education Research Service.

Hiebert, J. & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30*(2), 393 – 425.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givven, K. B., Garnier, H., Smith, M., Hollingsworth, H., Manaster, A., Wearne, D., & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Education Evaluation and Policy Analysis, 27*(2), 111 – 132.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Kazemi, E. & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal, 102*(1), 59 – 80.

Koichu, B., Berman, A., & Moore, M. (2007). The effect of promoting heuristic literacy on the mathematical aptitude of middle-school students. *International Journal of Mathematical Education in Science and Technology, 38*(1), 1 – 17.

Kosko, K. W., & Miyazaki, Y. (2009). Compound effects of mathematics discussion on fifth grade math achievement. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 284-287). Atlanta, GA: Georgia State University.

Kosko, K. W., & Miyazaki, Y. (in press). The effect of student discussion frequency on fifth graders' mathematics achievement in U.S. schools. *The Journal of Experimental Education*.

Lee, C. (2006). *Language for learning mathematics: Assessment for learning in practice.* New York: Open University Press.

Lopez, L. M., & Allel, L. (2007). Sociomathematical norms and the regulation of problem solving in classroom microcultures. *International Journal of Educational Research, 46*(5), 252-265.

Manouchehri, A. & St. John, D. (2006)From classroom discussions to group discourse. *Mathematics Teacher, 99*(8), 544 – 551.

McClain, K., & Cobb, P. (2001). An analysis of the development of sociomathematical norms in one first-grade classroom. *Journal for Research in Mathematics Education, 32*(3), 236-266.

Mercer, N. (2000). *Words & minds: How we use language to think together.* London: Routledge.

Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education, 20*(6), 507 – 528.

National Center for Education Statistics [NCES]. (2007). *Spring 2007 grade 8 math teacher questionnaire*. Retrieved from http://nces.ed.gov/ECLS/pdf/eighthgrade/teachermath.pdf

National Center for Education Statistics [NCES]. (2004). *Spring 2004 fifth grade child-level questionnaire: Mathematics teacher*. Retrieved from http://nces.ed.gov/ECLS/pdf/thirdgrade/teachermath.pdf

National Center for Education Statistics [NCES]. (2002). *Spring 2002 teacher questionnaire*. Retrieved from http://nces.ed.gov/ECLS/pdf/thirdgrade/teachermath.pdf

National Center for Education Statistics [NCES]. (2000). *Spring 2000 teacher questionnaire*. Retrieved from http://nces.ed.gov/ECLS/pdf/ firstgrade/teachersABC.pdf

National Center for Education Statistics [NCES]. (1999). *Kindergarten teacher questionnaire: Spring*. Retrieved from http://nces.ed.gov/ECLS/pdf/kindergarten/springteachersABC.pdf

National Center for Education Statistics [NCES]. (2009). *Early childhood longitudinal study, kindergarten class of 1998-1999 (ECLS-K): Combined users' manual for the ECLS-K eighth-grade and K-8 full sample data files and electronic codebook*. Retrieved from http://nces.ed.gov/ecls/dataproducts.asp

Pimm, D. (1987). *Speaking mathematically: Communication in mathematics classrooms.* New York: Routledge & Kegan Paul.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2007). *HLM6.26d: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Schleppenback, M., Perry, M., Miller, K. L., Sims, L., & Ge, F. (2007). The answer is only the beginning: Extended discourse in Chinese and U.S. mathematics classrooms. *Journal of Educational Psychology, 99*(2), 380 – 396.

Sfard, A. (2007). When the rules of discourse change, but nobody tells you: Making sense of mathematics learning from a commognitive standpoint. *Journal of the Learning Sciences, 16*(4), 565 – 613.

Shouse, R. (2001). The impact of traditional and reform-style practices on students mathematical achievement. In T. Loveless (Ed.), *The Great Curriculum Debate: How Should We Teach Reading and Math?* Washington, D.C.: Brookings Institution Press.

Silver, E.A., Kilpatrick, J., & Schlesinger, B. (1990). *Thinking through mathematics: Fostering inquiry and communication in mathematics classrooms.* New York: College Entrance Examination Board.

Springer, G. T. & Dick, T. (2006). Making the right (discourse) moves: Facilitating discussions in the mathematics classroom. *Mathematics Teacher, 100*(2), 105 – 109.

Stapes, M. & Colonis, M. M. (2007). Making the most of mathematical discussions. *Mathematics Teacher, 101*(4), 257 – 261.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction. *Phi Delta Kappa, 79*(1), 14 – 21.

Truxaw, M. P. & DeFranco, T. C. (2007). Lessons from Mr. Larson: An inductive model of teaching for orchestrating discourse. *Mathematics Teacher, 101*(4), 268 – 272.

Williams, S. R. & Baxter, J. A. (1996). Dilemmas of discourse-oriented teaching in one middle school mathematics classroom. *The Elementary School Journal, 97*(1), 21 – 38.

Wood, T. (1999). Creating a context for argument in mathematics class. *Journal for Research in Mathematics Education, 30*(2), 171 – 191.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458 – 477.