

Student Attitudes and Achievement in Active Learning Calculus

Pablo A. Duran^a, Adam J. Castillo^b, Charity Watson^b, Edgar Fuller^b,
Geoff Potvin^b & Laird H. Kramer^b

^aDepartment of Mathematics and Applied Mathematics,
Virginia Commonwealth University, Richmond, VA,

^bSTEM Transformation Institute, Florida International University, Miami, FL

The present paper explores the relationship between attitudes towards mathematics (ATM) and achievement in college calculus in active learning (AL) and lecture-based (LB) classrooms. Previous work on this relationship has mainly been limited to LB instruction, neglecting the impact of innovative approaches such as AL. Less attention has been paid to the roles played in this relationship by gender, year in college, and initial ATM. Results from a sample of 535 undergrad students enrolled in 9 AL and 9 LB sections are presented. Data included ATMI surveys responses, final grades, and demographics. Correlation and multiple regression analyses were conducted. The influence of instruction on students with low ATM was also examined. Gender and year in college were the main demographic variables considered. Achievement in AL was found to be less dependent on initial ATM in terms of correlation. AL showed higher gains in grades than LB, when controlling for ATM and demographic variables. Effect sizes of AL instruction on grades of students with low ATM were larger than those of students with higher ATM. Furthermore, AL courses had a large effect size ($d=0.81$) on female students with lower ATM, confirming its role as a gender equalizer.

Keywords: Active learning; gender equalizer; student achievement; college calculus; competitiveness, student attitudes.

Subject classification codes: 97D40

This is an original manuscript of an article published by Taylor & Francis in International Journal of Mathematical Education in Science and Technology on 13 Dec 2022, available at: <https://doi.org/10.1080/0020739X.2022.2150902>.

Introduction

Although college calculus has continuously been confirmed as a central component of STEM careers (Bressoud et al., 2013), its steady presence in the curriculum has been linked to large reform efforts (Bressoud, 2020). Some of the main drivers of these reform efforts include students' persistent low achievement (Bressoud et al., 2015), poor conceptual understanding of core ideas, and inability to transfer these ideas to other contexts in STEM (Hallett, 2006). These efforts have primarily focused on changes in instructional practices, content selection to highlight core ideas, and the emphasis on the application of core ideas to different disciplines in STEM (Carreon et al., 2018; Rasmussen et al., 2014).

In terms of changes in instructional practices, special attention has been paid to classrooms that have implemented active learning (AL) approaches. These approaches have been characterized as a shift from teacher-centered, lecture-based classrooms to more student-centered classroom environments where students are actively engaged in meaningful learning activities and have an opportunity to reflect on their learning process (Bonwell & Eison, 1991). Examples of these approaches include flipped classrooms (DeLozier & Rhodes, 2017; Jungić et al., 2015), inquiry-based learning (Kogan & Laursen, 2014; Laursen et al., 2014), and courses that incorporate a larger component of group work supported by mathematical software (Armstrong & Hendrix, 1999).

The effectiveness of active learning (AL) approaches started to appear in the literature over three decades ago (Davidson, 1971). Since then, it has been extensively investigated in undergraduate STEM education (Freeman et al., 2014; Johnson & Johnson, 2002). For instance, Freeman et al. (2014) in a meta-analysis of 225 studies, found that these approaches led to an increase in student performance on examinations and concept inventories by almost one-half standard deviation over traditional lecturing. Furthermore, students enrolled in courses based on these approaches were found to be 1.5 times less likely to fail the course

than traditional lecturing courses. This success has also proven to be consistent across STEM areas, leading to high odds of student success in biology (Chambers, 2008), chemistry (Paulson, 1999), and computer sciences (Lasserre, 2009). In other STEM disciplines, such as modeling in physics, these odds have shown to be as high as 6.73 times greater than lecture-based instruction (Brewer et al., 2010).

In the specific case of college calculus, current studies on the effectiveness of AL, however, are still limited. These studies have focused on high-stakes assessments and concept inventories, neglecting other relevant outcomes such as students' attitudinal changes. Furthermore, meta-analysis studies on the impact of AL have included only a small sample of calculus courses with a wide variety of AL strategies, from courses with a high-lecture component and marginal emphasis on student collaboration (Lindaman, 2007; Maggelakis & Lutzer, 2007), to courses with minimal lecturing in classes and large computer-based component (Armstrong & Hendrix, 1999). Further research on the effectiveness of AL on additional measures such as student attitudes in calculus is still needed.

As an additional measure of the impact of AL on students' learning process, attitudes towards mathematics (ATM) in calculus has only recently been examined (Castillo et al., 2022). Developing a better understanding of this influence is important because Calculus has been found to produce sharp declines in ATM in college (Bressoud, 2015). These declines can have dire consequences in students' experiences, impacting students' persistence in STEM programs (Bressoud et al., 2013; Maltese & Tai, 2011), classroom equity (Ellis et al., 2016), and driving subsequent mathematics avoidance (Popham, 2005). Finding evidence of a direct link between AL and ATM positive gains would provide further support for the implementation of this and similar instructional strategies in calculus, contributing to addressing these persistent issues in STEM education. The existence of this evidence in the literature remains extremely limited.

The influence of ATM on achievement, on the other hand, has been widely ascertained in postsecondary settings, finding small-to-moderate effect sizes in calculus (House, 1995; Pyzdrowski et al., 2013; Sonnert et al., 2020). This effect size has been translated, when controlling for other relevant predictors of achievement such as mathematics preparation, to a 10% increase in final grade per 1-SD increase in ATM (Pyzdrowski et al., 2013). The nature of the close interaction between students' ATM and their achievement could be associated with the existence of a feedback loop between these factors (Ma & Kishor, 1997; Sonnert et al., 2020). As the semester progresses, students' higher achievement in preliminary assessment increases their ATM, and this renewed ATM, in turn, helps them to increase their future achievement in the class. Conversely, students' lower achievement in preliminary assessment decreases their ATM, and such change undermines their future achievement. This reinforcement cycle aligns with the influence that performance accomplishments can have over students' self-efficacy. As Bandura suggests, students' "successes raise mastery expectations, repeated failures lower them." (1977, p. 195)

It is possible that the enhanced student engagement in AL classrooms could change the way in which this cycle evolves throughout the semester. The low-stake environment of AL classrooms and emphasis on student collaboration could contribute to a better distribution of skills, attitudes, and knowledge in the classroom (Slavin, 1996). Students with initially low ATM, often conditioned to lower achievement, might be positively influenced by students on the other extreme of the spectrum. Despite how plausible this mechanism seems to be, no studies were found that confirm its existence in AL classrooms in college calculus.

Further reasons supporting AL influence on ATM are related to some of the central elements present in some AL approaches, including cooperative learning and social metacognition. On the one hand, several psychological theories confirm the influence of cooperative learning on attitudinal development (Johnson & Johnson, 1991). As this type of

learning is characterized by students working together to achieve joint goals while cooperating in group discussions, students not only commit to adopt certain attitudes, but also advocate attitudes to others (Johnson et al., 2007).

On the other hand, although evidence of the influence of social metacognition on student attitudes is limited, there are reasons to expect a significant supportive role. Social metacognition can be understood as an extension of metacognition into group interactions, where peers influence awareness of one another's knowledge, emotions, and actions (Chiu & Kuo, 2010). Increases in this awareness could be strengthening students' confidence in their ability to be successful. Furthermore, a recent study in physics reported a positive correlation between metacognition and students' enthusiasm and self-confidence (Eblen-Zayas, 2016).

Additionally, AL courses that follow culturally appropriate learning models could even have a higher impact on students' attitudes. Some of the key features of these models include the creation of a safe, inclusive, and respectful learning environment and promotion of equitable learning (Ginsberg & Wlodkowski, 2009). In this environment students would feel more comfortable helping and giving evaluative feedback to peers, which are known elements of social persuasion hypothesized to directly influence attitudes (Bandura, 1977).

Equally important, in investigating the nature of the mechanism through which AL might change the relationship between initial attitudes and achievement, is the role demographic variables might be playing in this process. Clear interactions between ATM, achievement, and demographic variables, such as gender, year in college, career choice, race and ethnicity have been found. First, a recent study reported that women start and end the semester with significantly lower mathematical confidence than men, even when looking only at students with above-average mathematics preparation and skills (Ellis et al., 2016). Second, in terms of year in college, freshman students in college calculus have shown to have a more positive ATM than their sophomore counterparts (Sonnert et al., 2020). In terms of

career choice, students who were interested in pursuing STEM careers had higher ATM than the rest of the students enrolled in the course (Sonnert et al., 2020). Last, literature on the influence of race and ethnicity on ATM is much more limited, but a previous study reported evidence that African American students showed higher values of ATM than their White counterparts, while the opposite relationship was true when examining their achievement (Stanic & Hart, 1995).

Another critical aspect regarding the influence of AL in the classroom is its impact on students with particularly low attitudes. A regression model from a recent study suggests that students with initially low ATM (those with less than 2 standard deviations in mean ATM scores) are likely to earn low grades, fail or drop the class (Sonnert et al., 2020). There is evidence that these students benefit more from conventional good teaching with less group work than other instructional strategies better aligned with AL (Sonnert et al., 2015). However, a recent study found that certain AL approaches could lead to better outcomes in students' ATM (Castillo et al., 2022). This study however did not examine achievement, suggesting the need further research.

The main purpose of the present study is to examine the relationship between attitudes towards mathematics (ATM) and achievement in AL and LB calculus classrooms. The following five research questions guided this investigation: (1) To what extent are students' initial ATM associated with their achievement in both LB and AL classrooms? (2) How does this association vary by demographics? (3) To what extent do initial ATM predict achievement when controlling for instructional strategy and students' demographics including gender, year in college, and STEM choice? (4) To what extent does AL influence the achievement of students with initially low ATM differently, when compared to traditional classrooms? (5) How does this difference vary by demographics?

Related Literature

Attitudes Towards Mathematics

Extensive research on student affect has been conducted in mathematics over the last decades (Zan et al., 2006; Reyes, 1984). The general idea of how student attitudes towards mathematics (ATM) are portrayed in these reviews is summarized by Reyes as “students' feelings about mathematics, aspects of the classroom, or about themselves as learners of mathematics” (1984, p.1).

Understanding the impact of these feelings in the learning process is important as they have been linked to student achievement (Evans, 2007; Yee, 2010), and have also proven to play a predominant role in shaping students' persistence in STEM programs (Bressoud et al., 2013; Maltese & Tai, 2011). Furthermore, students' ATM have been recently related to issues of equity. A study examining the effect of attitudes by gender, for instance, found that odds of a female student being discouraged from continuing in calculus is 1.5 times greater than that for a male student (Ellis et al., 2016).

Due to the complexity of students' feelings, developing reliable instruments that measure students' ATM has proved to be a long process (Chamberlin, 2010; McLeod, 1994). This process started over six decades ago undergoing multiple iterations focused on capturing the most salient traits of student attitudes (Hannula, 2002). Most recently, the attitudes toward mathematics inventory (ATMI) was developed to capture the most essential dimensions of student attitudes but with less items than previous instruments (Lim & Chapman, 2013; Tapia & Marsh, 2004). It is composed of 40 items and four factors: confidence, enjoyment, motivation, and value. Moreover, its psychometric properties have been confirmed in a college setting, yielding good model fit statistics and high Cronbach alpha coefficients (above 0.87) for each factor (Tapia & Marsh, 2002) in college settings.

Influence of AL on Student Attitudes

Fewer studies have been conducted to understand the impact of AL approaches on students' attitudes than those examining students' achievement. Moreover, most of these studies have focused on pre-tertiary education. However, results in general suggest a positive impact on students' attitudes. In the US, findings from a meta-analysis of 65 studies on primary and secondary school (E. Savelsbergh et al., 2016) provide some evidence of this impact.

According to this study, significant effects ($0.35 < d < 0.4$) were found in general attitude, general interest, and career interest in science. These results included a variety of innovative instructional strategies such as inquiry-based, computer-based, and collaborative learning strategies with no significant difference between approaches. Another interesting outcome of this study is that the effects of instruction were found to be weaker for older students.

Outside of the US, the positive influence of AL on student attitudes has also been confirmed in secondary science education. For instance, in a study of over 12,000 students in the UK, student engagement in high school classrooms was linked to positive attitudes towards science, and particularly higher levels of student enjoyment and motivation (Hampden-Thompson & Bennett, 2013). Similarly, studies recently conducted with students in Turkey also found significant effect sizes on the impact of AL in students' attitudes (Akinoglu & Tandogan, 2007; Demirci, 2017).

In college calculus, Alkhateeb (2002) found that adding a hands-on, technology-based component to the course further enhanced students' attitudes and significantly reduced students' mathematics anxiety. Finally, a more recent study (Castillo et al., 2022) noted that incorporating AL strategies in a calculus course can improve student attitudes gains, when compared to a traditional lecture-based course. Moreover, this study found that AL had a particular positive impact on female students' self-confidence, acting as a gender equalizer.

Although there still is a need for more systematic studies, experiences in multiple educational settings in pre-tertiary education and STEM disciplines in college points to a positive influence of AL approaches on student attitudes.

With respect to how AL interacts with demographics variables including gender, year in college, and STEM intended choice a few, yet limited, studies suggest the existence of significant interactions. For example, Laursen et al. (2014) conducted a quasi-experimental study of over 100 course sections at four academic institutions over a period of two years to understand differences between inquiry-based learning (IBL) courses and comparable non-IBL courses. Their results indicated not only an overall positive impact of IBL, but a significant decrease in attitudinal and achievement gaps. While female students in non-IBL courses still showed much lower cognitive and attitudinal gains than their male counterparts, in IBL courses, no significant differences in these gains were found.

Research on the influence of AL on other demographics variables such as year in college, and STEM intended career choice is extremely limited. Only one study, by Fuselier and Jackson (2010) was found regarding this issue. This study examined students' views on how collaborative science might change depending on their coursework, finding that the fewer science courses students take, the more collaborative they report science is for them. This progression in students' views of science could influence the impact of AL classrooms by year in college, indirectly suggesting that AL might be more effective in students in their first years of college. The present study is intended to contribute to the limited evidence found. Before presenting main findings, details of the methodology used are included as follows.

Methods

The present study investigated the influence of AL on the relationship between students' initial attitudes towards mathematics and their achievement on an introductory calculus

course at a large, urban, research intensive (R1) university. It initially included a randomized control trial experiment during the Spring 2019 and Fall 2019 terms to establish strong and reliable evidence. Instructors in the treatment group used an AL curriculum following the Modeling Practices in Calculus (MPC) model. Instructors in the control group, on the other hand, followed a lecture-based model.

Modeling Practices in Calculus

Faculty who followed the MPC model, participated in professional development that included a three-day workshop prior to teaching, and weekly planning meetings throughout the semester to support the model adoption. The MPC model integrates three core elements at its foundation: cooperative learning (Johnson et al., 2007; Johnson & Johnson, 1991), social metacognition (Chiu & Kuo, 2010), and a culturally appropriate learning environment (Ginsberg and Wlodkowski, 2009). First, in terms of cooperative learning, students work most of the class in groups on a set of notes and learning activities that develop student understanding of core calculus ideas. The notes introduce the key topics of the day with some examples and questions for the groups, and the learning activities contain a set of problems that lead students to reflect on and challenge their understanding of these topics. For example, after just the second class, students are led through the notes through group and whole-class discussions to have an intuitive understanding of limits using visual representations. With these notes, groups are then asked to develop an idea of how to compute limits without the need for a graphical or tabular representation. The associated learning activity presents a question involving a piecewise-function and asks to compute limits without any visual representations. By working cooperatively on the learning activity, students note how they use the domain of the piecewise function to ‘visualize’ what happens which helps them build an understanding of direct substitution, the first technique for computing limits presented.

Second, the MPC model includes social metacognition as an essential element of the class. Opportunities for developing social metacognition are promoted on a typical day of class as students work together to write up and present ideas and solutions to problems they developed in their groups on whiteboards. Group members are asked to monitor each other's thinking and make suggestions to control their group problem solving.

Last, MPC's culturally appropriate learning model allows students to try out their ideas in a low-stakes, safe environment, receive ongoing formative feedback from an instructional team, and participate in a community of learners. The instructor promotes a safe learning environment by messaging to students regularly that making mistakes and asking questions are acceptable and a natural part of mathematics. The low-stakes environment is also enhanced as Learning Assistants (LAs), or trained undergraduate classroom facilitators, are integrated into the classroom to support learning with groups and provide valuable information to instructors about student interactions (Otero et al., 2010). LAs are natural agents of this culturally appropriate model, as their demographics are that of the students, who provide insights and connections from the point of view of a former student in the course.

Participants

The sample in this study consisted of a total of 553 students enrolled in Calculus I at a large, urban, research-intensive institution in the US. In the Spring 2019 semester, a total of 168 of these students were randomly assigned to three control and three treatment sections. In the subsequent semester, the number of sections increased due to semester enrollment trends and the gradual AL curriculum implementation design. In the Fall 2019 semester, the total of students participating in this study then expanded to a total of 385 students randomly assigned to six control and six treatment sections.

Additionally, students' demographics data was reported by students to the university and collected at the time of course enrollment. A breakdown of the number of students by treatment group and the demographics of all participating students, can be seen in Table 1.

Table 1. Demographics by treatment group

	Treatment (N=286)	Control (N=267)		Treatment (N=286)	Control (N=267)
<i>Gender</i>			<i>STEM choice</i>		
Female	147 (51.4)	117 (43.8)	Non-STEM	50 (17.5)	45 (16.9)
Male	119 (41.6)	132 (49.4)	STEM	216 (75.5)	204 (76.4)
Missing/NA	20 (7.0)	18 (6.7)	Missing/NA	20 (7.0)	18 (6.7)
<i>Precalculus proficiency</i>			<i>Race/Ethnicity</i>		
High	75 (26.2)	78 (29.2)	African Am.	15 (5.2)	16 (6.0)
Low	186 (65.0)	164 (61.4)	Asian/Pac. I.	12 (4.2)	8 (3.0)
Missing/NA	25 (8.7)	25 (9.4)	Hispanic	185 (64.7)	199 (74.5)
<i>Class Standing (Year in College)</i>			White	32 (11.2)	15 (5.6)
Freshman	93 (32.5)	79 (29.6)	Others	22 (7.7)	11 (4.1)
Sophomore	86 (30.1)	93 (34.8)	Missing/NA	20 (7.0)	18 (6.7)
Junior	60 (21.0)	49 (18.4)			
Senior	27 (9.4)	27 (10.1)			
Others/NA	20 (7.0)	19 (7.1)			

Note. In parentheses: percentages of students in each category in both semesters.

Measures and Procedure

Measures

Student attitudes were measured using the Attitudes towards Mathematics Inventory (ATMI) developed by Tapia and Marsh (2004). This survey is composed of 40 items measuring the four subscales described in the previous section of this study: enjoyment (10 items), motivation (5 items), self-confidence (15 items), and value (10 items). Eleven items

of this survey were reversed-coded later on for data analysis. Additionally, at the end of the semester, students' final grades were collected and converted to a 100-point scale (A+ = 98, A = 94.5, A- = 92, B+ = 88, B = 84.5, B- = 81, C+ = 78, C = 74.5, C- = 71, D+ = 68, D = 64.5, D- = 61, F = 40). This scale was previously used in a study (Sonnert et al., 2015) on calculus performance that included student attitudes as a predictor, yet not accounting for innovative instructional strategies as active learning.

Procedures

In order to obtain the final sample in this study, students enrolled in multiple, 80-seat (twice the normal size) sections of introductory calculus, chosen to fit their schedules as they normally would. Instructor names were invisible to students throughout this enrollment process. Two days prior to the beginning of each term, each of these 80-seat sections were then split into two 40-seat sections by assigning each student at random to one of either a treatment (MPC) or control (non-MPC lecture-based traditional instruction) section. After this random assignment was completed, students were still allowed to change sections prior to the enrollment deadline.

In the Spring semester, a total of 261 students were randomly assigned to ten sections within the study, with 130 students of these in five treatment sections and 131 students in five control sections. In the Fall 2019 Semester, a total of 533 students were randomly assigned to 16 sections within the study, with 271 students in eight treatment sections and 258 students in eight control sections. Since only sections with matching schedules (same day/time teaching) were included in this study, the final sample included three sections per treatment in the Spring 2019, and six sections per treatment in the Fall 2019 semester. A group of students, no larger than 21% of the sample, were not part of the original random assignment. This group included students from other sections who decided to enroll in RCT sections after the split and prior to the enrollment deadline. However, these students enrolled evenly in control

and treatment sections. The number of students per treatment per semester in the final sample can be seen in Table 1. Students were asked to complete the ATMI and the PCA survey, at the beginning (first week of classes) and end of the semester (last two weeks of classes). Surveys were administered by the instructors, following a protocol that involved ensuring students their participation was not going to influence their grade in any way.

Missingness

The overall unweighted unit response rate for both semesters was 82.7% for pre-surveys, 66.8% for post surveys, and 64.8% for students submitting both surveys. A breakdown of survey response rates by semester is presented in Table 2.

Table 2. Initial enrollment and survey-response rates by treatment section

	Control	Treatment
Enrollment ^a	335	342
Pre-survey ^b	248 (74.0)	263 (76.9)
Post-survey ^b	178 (53.1)	235 (68.7)
Both surveys ^b	156 (46.6)	204 (59.6)

^a End of the semester enrollment count, ^b Number of students answering at least one item of the respective survey (%)

A total of 82 (8.8%) out of 924 pre- and post-surveys collected were partially completed (at least one item completed). Excluding blank surveys, item non-response rate for both treatment and control groups for each item was less than 3.4% for both pre- and post-surveys. When including blank pre-ATMI surveys later paired with partially completed post-surveys, on the other hand, the percentage of missing values across the 40 pre-survey items of collected surveys ranged between 9% to 14%. However, this percentage was much higher for the post-survey items, ranging between 18% to 37%. This unbalance was mainly explained by high rates of student attrition, especially in the control section.

Due to differences in attrition rates between control and treatment sections, missing data was considered missing not completely at random (Rubin, 1975). Potential loss of statistical power and biased estimates due to this missingness were addressed using a multiple imputation (MI) algorithm (Kang, 2013). The extensively validated expectation-maximization with bootstrapping (EMB) MI algorithm AMELIA II was used to impute unit and item non-response (Honaker et al., 2011). MI was carried out considering the pre-post design as a time series, using the pre PCA survey results as a covariate. Since the percentage of missing data was less than 30%, using over 30 iterations was considered appropriate (White et al., 2011). Although assumptions of normality were violated for each survey item, given the sample size in this study and since these deviations were slight, the EMB algorithm was assumed to be robust against these violations (Demirtas et al., 2008).

Data Analysis

Four different strategies were used to analyze the data: a correlation analysis, a least square dummy variable (LSDV) fixed effects model, an optimal cutoff analysis, and an analysis of mean differences for students identified with low and high initial ATM. First, a correlational analysis was carried out to understand the strength of association between ATM and achievement overall and by demographics. Student achievement and ATMI scores were both considered continuous interval variables. Pearson product moment correlation coefficient preferred, since sample sizes were considered large enough to assume normal distribution (Bujang & Baharum, 2016). Since missingness was addressed using MI, correlation coefficients were pooled using Fisher's Z transformation to normalize data before using Rubin's rules (Enders, 2010, p. 220).

Second, it was expected that instructors differences might have led to slight variations in MPC implementation . The lack of direct measures of fidelity of implementation is a limitation of this study, however, as previously indicated, weekly supporting meetings with

instructors, and professional development workshops contributed to control for this issue. Additionally, a least square dummy variable (LSDV) fixed effects model, preferred over mixed-effect models (McNeish & Kelley, 2019; McNeish & Stapleton, 2016), accounted for differences in sections.

The response variable was students' final grades (100-point scale), the only continuous predictor was students' initial ATMI scores, the remaining predictors were all categorical: treatment (0:Non-MPC sections; 1: MPC sections), gender (0: male students, 1: female students), year in college (0: freshmen, 1: sophomore, 3: others), and STEMdeclared (0: Non-STEM and 1: STEM). The section variable included to account for differences in instructors had 12 levels in the MPC sections and 12 levels in the non-MPC sections.

Following findings reported in the literature review (Sonnert et al., 2015), interactions between treatment and initial ATM were added to the model. Additional interactions were also included between treatment and each of the following variables: gender, year in college, and STEMdeclared. Although evidence of these additional interactions was not found, each variable was expected to have large main effects, thus likely to interact with treatment. Additionally, their inclusion was expected to expand the understanding of the relationship among these variables (Harrell, 2001). In terms of variable selection, since this study was exploratory, and there were a priori reasons to assume the initial variables were all relevant, the full model was preferred over stepwise methods. Furthermore, stepwise methods were also avoided to minimize bias in parameter estimation, and error inflation (Harrell, 2001).

Third, from previous studies the effect of ATM on achievement was expected to be polarized (Sonnert et al., 2015). An optimal cutpoint analysis using the Youden index metric (Yin & Tian, 2014) was conducted on the control section as baseline to identify groups of students with low and high attitudes. This cutoff optimized the likeliness of misclassifying

students' success based on their initial ATM. Students with initial ATMI scores lower than this cutoff were considered more likely to fail the class, based on the data collected from the Non-MPC sections. Results of this analysis were also compared to the cutoff score analysis based on equal-frequency discretization to check for consistency. The optimal cutpoint analysis was conducted using the `cutpointr` R package (Thiele & Hirschfeld, 2021).

Last, after identifying groups with low and high initial attitudes based on the previously determined optimal cutoff score, the ATMI scale was considered a continuous interval variable, given the Likert scale nature of the items and large sample size of this study (Carifio & Perla, 2008). This assumption followed previous studies (Asante, 2012; Karjanto, 2017; Primi et al., 2020) and was supported by two main theoretical positions. On one hand, the ATMI scale was considered to have no true zero. Each item in the ATMI survey was scored with a minimum of 1 (strongly disagree) and a maximum of 5 (strongly agree). A score of 3, corresponding to the "Neutral" response, was also considered a student attitude and not the lack of it. When adding each item's score, the overall ATMI scale ranged accordingly between 40 and 200.

On the other hand, the assumption of equal distance between points was deemed to be reasonable given the type of Likert scale involved (Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree). Furthermore, confirmatory analyses on the ATMI survey (Lim & Chapman, 2013; Ngurah & Lynch, 2013) suggest that the contribution of each item to each scale (overall, motivation, enjoyment, self-confidence, and value scale) is fairly homogeneous. Since Likert items on each scale in this study were not examined individually, but as summated scales, this homogeneity prevented certain items from over or under representation of the scale. Therefore, 1-point differences in scores were assumed to be similar.

Limitations associated with the midpoint choice of neutral response (Chyung et al., 2017) and the presence of small differences in loadings from item to item in confirmatory analyses (León-Mantero et al., 2020) were assumed to be controlled by the robustness of the summated scales, given the sample size, the number of points in each item, and the number of items in each scale in this study (Carifio & Perla, 2008; Pell, 2005). Finally, multiple ANCOVAS were then conducted separately in two groups: low and high initial ATM. Results were used to compare students' achievement between MPC and Non-MPC sections, while controlling for initial ATMI scores, and semester of enrollment. Since the nature of this study was exploratory, corrections for multiplicity were not conducted to prevent error type II inflation (Streiner, 2015). All statistical analysis, unless otherwise indicated, were conducted in R, using the stats package (v4.0.2; R Core Team 2021).

Results

Correlation Analysis

Pearson product-moment correlation coefficients were computed to measure the strength of relationship between student attitudes towards mathematics, as measured by the ATMI scores, at the beginning of the semester and their course achievement, as measured by their total grade at the end of the semester. Correlation coefficients were pooled from the multiple imputed datasets using Fisher's Z transformation and Rubin's rules. Results of this analysis are presented in Table 3.

Table 3. Pearson's pooled correlation coefficients of ATMI pre-scores and final grade by demographics

Group	df	MPC ^a	df	Non-MPC ^a
<i>Overall</i>	265	0.193	248	0.238
<i>Gender</i>				
Female	146	0.135	116	0.252

Male	118	0.274	131	0.260
1-2 Years	162	0.118	150	0.224
>2 Years	79	0.340	65	0.247
<i>STEM declared</i>				
STEM	215	0.188	203	0.270
Non-STEM	49	0.244	44	0.130

Note: ^a Pearson's correlations coefficients pooled using Fisher's Z transformation and Rubin's rules. df=N-1. All coefficients were statistically significant ($p < .001$) ^b based on a cutoff of 60 credits.

As shown in Table 3, correlation between student attitudes and achievement in both MPC and Non-MPC sections was found to be small-to-moderate (Cohen, 1988). Students' attitudes in MPC sections were slightly less correlated to achievement. When considering specific groups by gender, year in college, and STEM declared. When examining gender, attitudes were significantly less correlated to achievement for MPC female students than in non-MPC sections. On the other hand, correlations for male students were similar in both MPC and Non-MPC sections.

Regarding year in college, attitudes were correlated to achievement in a similar way for freshman and sophomore students in both types of instruction with significantly less correlation in MPC students. In senior students, on the other hand, this situation was reversed. Finally, correlation between students' attitudes and achievement was mixed for STEM and Non-STEM students. Correlations were significantly lower for MPC sections in STEM students, but higher in non-STEM students. Although most correlations were found to be small-to-moderate, differences in the strength of association by treatment confirmed the need to further investigate these relationships using a regression analysis.

Multiple Regression Analysis

A least square dummy variable (LSDV) fixed effects model was used to predict students' final grade based on students' ATMI scores at the beginning of the semester, type

of instruction (MPC or Traditional), gender (female or male) and year of college (freshman, sophomore, junior or senior). The Section variable was considered a fixed effect to control for between-variability due to differences between instructors or groups of students who might attend different class days. Following a recent study (Sonnert et al., 2020) absence of non-linearity in the relationship between initial ATM and Grade was assumed. Basic descriptive statistics and regression coefficients of this model are shown in Table 4.

Table 4. Descriptive statistics and regression coefficients of LDSV model

	estimate ^a	std.error	t	conf.low	conf.high
(Intercept)	78.24	3.22***	24.33	71.92	84.56
ATMI	3.23	1.02**	3.18	1.23	5.23
ATMI:MPC	-0.52	1.33	-0.39	-3.13	2.09
ClassOthers	-4.60	2.35.	-1.96	-9.22	0.01
ClassSophomore	-4.89	2.16*	-2.26	-9.13	-0.64
GenderFemale	2.74	1.83	1.50	-0.85	6.33
STEMSTEM	4.54	2.36.	1.93	-0.09	9.17
MPC	9.82	5.69.	1.73	-1.36	21.00
MPC:ClassOthers	-1.22	3.19	-0.38	-7.49	5.06
MPC:ClassSophomore	0.22	3.00	0.07	-5.68	6.12
MPC:GenderFemale	-0.89	2.49	-0.36	-5.79	4.00
MPC:STEMdeclaredSTEM	-3.58	3.28	-1.09	-10.03	2.87

Note. ^a ATMI scores are normalized by z-scores. Signif. Codes: . p<.1, * p<0.05, ** p<0.01, *** p<0.001

For the initial model, about 11% of the variance in final grades were accounted for by the model (adjusted $r^2 = 0.107$). Such a low percentage was partially explained by excluding other relevant predictors of achievement such as mathematics preparation or students' sense

of belonging. ATMI initial scores were found to have a significant effect on Grade ($p < 0.05$), leading to an increase of 1-SD in ATMI initial scores per 3.23-point mean increase ($SE = 1.02$) in final grade in non-MPC sections. This increase in grade was found to be similar to the 3.15-point increase reported in a previous nationwide study (Sonnert et al., 2020).

Since the year in college factor had four different levels, an omnibus F test was conducted, finding a significant effect ($F = 8.7265$; $p < 0.0002$). None of the interactions included in the model (MPC interaction with gender, year in college, and STEMdeclared) were significant.

It is worth noting that the increase in grade in the model was found to be not significant, when controlling for all variables. Given the large standard error of the predictor, this result does not imply that AL had no impact on students' grade, but that follow-up studies are needed to better understand this impact. The model also failed to reflect the differences in gender and intended career choice previously reported in the correlation analysis.

When visually inspecting residual plots, the existence of a heavy-tailed distribution for the dependent variable was identified. Although estimation of model coefficients was robust against deviations from normality when sample size is reasonably large (Pek et al., 2018), this distribution could imply that data points in the tails were excessively penalized. The extreme value analysis reported in the following section was expected to contribute to measuring the impact of this penalization.

Achievement in lower ATMI scores

Optimal Cut-off score

In order to identify groups of students with extreme values of initial ATM, an optimal cut point analysis using the Youden index as a metric was conducted on the control section as

a baseline to identify students likely to pass the course as predicted by their initial ATMI score. A maximum value of 0.25 for the Youden index was achieved by setting the cut-off score for a passing grade in non-MPC sections at a raw ATMI initial score of 131 (sensitivity 69.6%, specificity 55.8%), corresponding to a mean score of 3.28 and a z-score of -0.28. Low correlation between ATMI scores and achievement led to a relatively low (AUC 0.62) accuracy of the cutoff in discriminating between students who pass or do not pass the class. However, the cutoff captured a high proportion of students with lower grades in both sections. This cutoff also seems to appropriately differentiate students earning higher grades (A or B), implying an extremely low probability of earning these grades with an initial ATMI score lower than this cutoff. Furthermore, equal-frequency discretization analysis led to similar results.

It is important to notice that this cutoff score should not be used to primarily predict students' performance. The sensitivity of the cutoff score clearly discourages such interpretation, as it estimates that this score would misclassify 30% of students with low ATMI scores as failing the class in the control sample. The cutoff score in this study was used only as an optimal score given the control sample data to identify groups of students with extreme values of ATM.

Differences by gender and year of college for students with low ATM

After identifying groups of students with lower initial ATM in the control sections, multiple two-tailed unpaired t-tests were conducted to investigate differences in final grade between MPC and Non-MPC sections. Using the previously found cut-off score (ATMI<131), results were summarized in Table 5.

Table 5. Final grade means, standard errors and effect sizes between MPC and Non-MPC sections by demographics for students under the optimal cutoff score (ATMI<131)

MPC	Non-MPC	Mean Differences
-----	---------	------------------

	n ₁	M	SD	n ₂	M	SD	t ^a	Δ	SE	LCI	UCI	d
<i>Low ATM</i>												
Female	45	80.72	12.45	46	73.71	12.17	2.74*	7.01	2.56	1.91	12.10	0.81
Male	35	76.82	14.34	30	71.64	18.45	1.22	5.18	4.24	-3.34	13.70	0.45
Freshm/Soph	45	80.81	14.12	44	70.68	15.94	3.2*	10.13	3.17	3.82	16.44	0.96
Jun/Senior	35	76.36	12.39	32	75.94	12.85	0.13	0.41	3.16	-5.92	6.75	0.05
<i>High ATM</i>												
Female	88	83.01	13.87	59	82.04	11.98	0.45	0.97	2.16	-3.30	5.25	0.12
Male	75	83.99	9.25	82	79.26	14.95	2.41*	4.73	1.96	0.85	8.61	0.53
Freshm/Soph	118	83.91	12.34	107	81.96	12.69	1.16	1.95	1.67	-1.35	5.25	0.23
Jun/Senior	45	82.40	11.22	34	75.74	15.44	2.12.	6.65	3.14	0.36	12.95	0.73

Note. ^aSignif. Codes: . p<.1, * p<0.05, ** p<0.01, *** p<0.001

As can clearly be seen in Table 5, AL did not negatively impact any of the groups considered in the low or high ATM categories. Furthermore, in the low ATM category the effect of AL on achievement was particularly larger for two groups: female students, with a medium-to-large effect size (Cohen's $d = 0.81$); and students in their first two years of college with a large effect size (Cohen's $d = 0.96$). The only group where the effect of AL on achievement was not significant in this category was students in their last years of college, with a negligible effect size (Cohen's $d = 0.05$).

It is also worth noting that the positive impact of AL in the achievement of students with high ATM was particularly higher for male students, with a medium effect size (Cohen's $d = 0.53$); and students in their last years of college, with a medium-to-large effect size (Cohen's $d = 0.73$). Some of the underlying reasons that could explain these trends are discussed in the following section.

Discussion

The main objective of this study was to investigate the relationship between attitudes towards mathematics (ATM) and achievement in two different college calculus settings: active learning (AL) and lecture-based (LB) classrooms. Previous work on this relationship has mainly been limited to LB instruction, and paid little attention to the roles played by gender, year in college, and low initial ATM. This study was intended to contribute to a better understanding of these issues through the results from an initial RCT design with a

sample of 535 students enrolled in control and treatment sections during the fall and spring semester of 2019. Treatment sections adopted the Modeling Practices in Calculus (MPC) approach which incorporates AL strategies and enhances learning facilitation by Learning Assistants (Otero et al., 2010), and control sections were predominantly lecture-based (LB) classrooms. Data collected from this implementation included a measure of students' ATM, using the ATMI survey (Tapia, 1996; Tapia & Marsh, 2002), final grades, and certain demographics. After a multiple imputation algorithm was used to address data missingness issues, the analysis of this data consisted of a correlational analysis, a fixed-effect model, and a mean differences analysis of students with low ATM, using a referential cutoff score found by metric optimization. A brief discussion of this analysis is presented below.

Impact of initial ATM on Achievement in AL classrooms

Since AL approaches have proven to positively impact both student achievement (Freeman et al., 2014) and ATM (Castillo et al., 2022), it was reasonable to expect that AL would have a moderation effect on students' initial ATM and achievement. We suspected that the higher engagement that students typically have in AL classrooms with peers and instructors would provide them with a better support system to overcome the detrimental effect of low ATM on their learning process, and thus reinforce their achievement.

The small positive correlations found between students' initial ATM and their achievement in MPC and Non-MPC classrooms were within the range of values previously reported (House, 1995). A significant difference between both classroom settings was found, indicating a lower correlation in AL classrooms. The direction of this correlation suggests the existence of a supportive role, in which AL helps students with lower ATM to earn higher grades than students in LB classrooms. Lastly, when considering specific groups of students, the correlation analysis suggested that AL's supporting role is particularly enhanced for female students, and students in their first two years of college.

The regression model confirmed the supportive role of AL on students learning processes including a significant increase in final grade, when controlling for ATM, gender, year in college, and STEM intent. Although we expected to find significant interactions between the control variables, none were identified. These results do not necessarily contradict the correlational analysis findings, as the lack of significance of a t-test associated with a regression coefficient cannot be directly interpreted (Harrell, 2001). We believe that two main shortcomings related to the regression model might have hindered its ability to accurately capture these interactions.

On the one hand, the available data did not include important predictors of achievement including variables associated with mathematics preparation. This restriction could explain the model's low goodness-of-fit (adjusted $r^2 = 0.117$) that prevented the model from accurately reflecting relationships between its variables (Fan & Huang, 2001). On the other hand, the existence of a heavy-tailed response variable distribution could have led to underrepresenting the influence of AL for low values of ATM (Catoni, 2012). An extension of our study to a larger sample and richer dataset that include additional predictors of achievement will contribute to clarify this issue. While addressing the limitations associated with the regression model, results from our analysis of students with low ATM not only helped to confirm some of the trends identified in the correlation analysis, but also provided further insight into the role that students' demographics might be playing.

Active learning as leverage to low initial ATM

The mean difference analysis confirmed expectations of a supportive role of AL in students with low ATM (Sonnert et al., 2015). In terms of demographic variables, as shown in Table 5, the most compelling findings were the large effect sizes identified in gender and year in college for students with low initial ATM. Our analysis showed that well-designed AL strategies such as MPC can have a particularly large effect ($d=0.81$) on female students

with low initial ATM. This effect size seems to initially contradict a recent study by Sonnert et al. (2020) who found that ATM's influence on achievement was similar for both male and female students. However, the regression model used by these authors was not focused on active learning classrooms nor included interactions specifically measured on students with extreme values of ATM, as we did in our study.

The large effect size on MPC female students with low ATM found in our study translates to an increase of over half-a-letter on final grade, when compared to traditional lectures. Such a large increase is of particular significance, since several studies have shown that female students enter STEM careers in college with lower ATM than their male counterparts (Good et al., 2012; Saxe et al., 2015). Furthermore, in terms of persistence, female students' odds of being discouraged from continuing in calculus have been estimated to be 1.5 times greater than that of their male counterparts (Ellis et al., 2016).

The impact of AL on female students whose initial attitudes might otherwise prompt them to fail the class, could contribute to create more equitable Calculus classrooms in college. Results from a recent study (Laursen et al., 2014) that compared inquiry-based learning (IBL) college mathematics to lecture-based classrooms are consistent with the supporting role we found AL's classrooms are playing in female students. In this study, differences between female and male students found in mastery gain in lecture-based classrooms vanished in IBL classrooms.

The underlying mechanism that explains the influence of AL on female students has not been fully explained. However, it is possible that the enhanced collaboration in AL classrooms provides female students with support they normally do not encounter in more lecture-based classrooms. This explanation is aligned with several studies that have shown that women, in general, respond more favorably to collaborative than competitive environments (Ash et al., 2004; Niederle & Vesterlund, 2011). Studies that support this

explanation in calculus are still needed, but studies on college physics confirm that the collaborative components of AL courses might be the main factor responsible for diminishing gaps in conceptual understanding (Lorenzo et al., 2006), and some attitudinal variables such as students' self-efficacy (Espinosa et al., 2019).

Regarding the achievement of students with low ATM by year in college, as shown in Table 5, AL was clearly more beneficial to those in their first or second year of college. This result is aligned with findings by Savelsbergh et al. (2016) in which the effect of innovative instruction was found to be weaker for older students. Laursen et al. (2014), also found similar results, where first-year students had greater gains in social and cognitive measures in IBL college mathematics than last-year students. Reported gains by Laursen et al. included mathematical thinking, persistence in solving problems, and collaboration. Although the reasons why these gains were observed were also not fully explained in their study, we suspect that since AL classrooms are still underrepresented in the STEM curriculum, junior and senior students might have a more extended experience with lecture-based courses in STEM, and that this experience makes their transition to AL more difficult. It is also possible that students' different views by year in college regarding collaboration and competition in science (Fuselier & Jackson, 2010) might be interacting with the benefits of AL classrooms.

Limitations and Further Research

The main limitations of this study were related to its sample size and the existence of unaccounted variables. In the first place, a larger sample size would allow for higher statistical power in each bivariate analysis conducted for each demographic variable and alleviate issues related to deviation from normality of the ATM scale. Having a larger sample would also allow the use of other regression models that account for extreme values such as spline regression (Harrell, 2001). Similarly, including more participants in the study could have helped to address a limitation related to natural constraints of students' enrollment in

college calculus. Giving students the freedom to switch sections after the ATMI survey was administered, resulted in a small group of students who were not part of the original random assignment. Although the fact that these students enrolled evenly in AL and traditional calculus sections led to an overall balanced sample, a proximity score matching algorithm (PSM) (Granger et al., 2019) could have contributed to better control for these changes. The reduced number of observations after PSM would however prevent us from exploring issues related to students' demographics with appropriate power. Expanding our study to additional semesters and other institutions would also help to address this limitation.

The authors also recognize the need to incorporate additional variables in a subsequent model. Some of the important unaccounted variables in this study included student mathematics preparation, instructors' characteristics, and measures of student collaboration. First, measuring students' mathematics preparation would have helped to provide a more robust regression model to capture the main interactions investigated. Second, this study assumed that the instructors' weekly professional development and planning meetings in the AL sections were effective in ensuring a high level of fidelity of implementation, yet no associated measures to confirm this assumption were included in the study. To partially address this issue, the effect of the instructor on the interactions was examined through a fixed-effect model with a dummy variable on instructor and its analysis yielded consistent results. Last, other variables such as students' course engagement or their level of collaboration throughout the semester were not available. Qualitative studies that explore students' classroom interactions in groups are also needed to avoid overlooking these and other relevant variables in this process.

References

- Akinoglu, O., & Tandogan, R. (2007). The effects of problem-based active learning in science education on students' academic achievement, attitude and concept learning. *Eurasia Journal of Mathematics, Science & Technology Education*, 3(1), 71–81.
- Alkhateeb, H. (2002). A preliminary study of achievement, attitudes toward success in mathematics, and mathematics anxiety with technology-based instruction in brief calculus. *Psychological Reports*, 90, 47–57.
- Armstrong, G. M., & Hendrix, L. J. (1999). Does traditional or reformed calculus prepare students better for subsequent courses? A preliminary study. *Journal of Computers in Mathematics & Science Teaching*, 18(2), 95–103.
- Asante, K. O. (2012). Secondary students' attitudes towards mathematics. *IFE Psychologia : An International Journal*, 20(1), 121–133. <https://doi.org/10.10520/EJC38916>
- Ash, A. S., Carr, P. L., Goldstein, R., & Friedman, R. H. (2004). Compensation and advancement of women in academic medicine: Is there equity? *Annals of Internal Medicine*, 141(3), 205–212.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bonwell, C. C., & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom*. 1991 ASHE-ERIC Higher Education Reports. ERIC.
- Bressoud, D. M. (2020). The strange role of calculus in the United States. *ZDM*, 1–13.
- Bressoud, D. M., Carlson, M. P., Mesa, V., & Rasmussen, C. (2013). The calculus student: Insights from the Mathematical Association of America national study. *International Journal of Mathematical Education in Science and Technology*, 44(5), 685–698.
- Bressoud, D. M., Mesa, V., & Rasmussen, C. L. (2015). *Insights and recommendations from the MAA national study of college calculus*. MAA Press.
- Brewe, E., Sawtelle, V., Kramer, L. H., O'Brien, G. E., Rodriguez, I., & Pamelá, P. (2010). Toward equity through participation in Modeling Instruction in introductory university physics. *Physical Review Special Topics-Physics Education Research*, 6(1), 010106.
- Bujang, M., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World Journal of Social Science Research*, 3(1), 37–46.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42(12), 1150–1152.

- Carreon, F., DeBacker, S., Kessenich, P., Kubena, A., & LaRose, P. G. (2018). What is old is new again: A systemic approach to the challenges of calculus instruction. *PRIMUS*, 28(6), 476–507.
- Castillo, A. J., Durán, P., Fuller, E., Watson, C., Potvin, G., & Kramer, L. H. (2022). Student attitudes in an innovative active learning approach in calculus. *International Journal of Mathematical Education in Science and Technology*, 1-29.
- Catoni, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'IHP Probabilités et Statistiques*, 48(4), 1148–1185.
- Chamberlin, S. (2010). A review of instruments created to assess affect in mathematics. *Journal of Mathematics Education*, 3(1), 167–182.
- Chambers, S. L. K. (2008). *Improving student performance in an introductory biology majors course: A social action project in the scholarship of teaching* [Ph.D., Union Institute and University].
- Chiu, M. M., & Kuo, S. W. (2010). From metacognition to social metacognition: Similarities, differences, and learning. *Journal of Education Research*, 3(4), 321–338.
- Chyung, S. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement*, 56(10), 15–23.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn. Á/L).
- Davidson, N. (1971). The small group-discovery method as applied in Calculus instruction. *The American Mathematical Monthly*, 78(7), 789–791.
- DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: A review of key ideas and recommendations for practice. *Educational Psychology Review*, 29(1), 141–151.
- Demirci, C. (2017). The Effect of Active Learning Approach on Attitudes of 7th Grade Students. *International Journal of Instruction*, 10(4), 129–144.
- Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation*, 78(1), 69–84.
- Eblen-Zayas, M. (2016). The impact of metacognitive activities on student attitudes towards experimental physics. *Proceedings of the 2016 Physics Education Research Conference, Sacramento, CA*, 104–107.

- Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit. *PLOS ONE*, *11*(7), e0157447.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Espinosa, T., Miller, K., Araujo, I., & Mazur, E. (2019). Reducing the gender gap in students' physics self-efficacy in a team- and project-based introductory physics class. *Physical Review Physics Education Research*, *15*(1), 010132.
- Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, *17*(2).
- Fan, J., & Huang, L.-S. (2001). Goodness-of-fit tests for parametric regression. *Journal of the American Statistical Association*, *96*(454), 640–652.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.
- Fuselier, L., & Jackson, J. K. (2010). Perceptions of collaboration, equity and values in science among female and male college students. *Journal of Baltic Science Education*, *9*(2), 10.
- Ginsberg, M. B., & Wlodkowski, R. J. (2009). Professional learning to promote motivation and academic performance among diverse adults. *Learning Never Ends*, *23*.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, *102*(4), 700.
- Granger, E., Sergeant, J. C., & Lunt, M. (2019). Avoiding pitfalls when combining multiple imputation and propensity scores. *Statistics in Medicine*, *38*(26), 5120–5132.
- Hallett, D. H. (2006). What have we learned from calculus reform? The road to conceptual understanding. *MAA NOTES*, *69*, 43.
- Hampden-Thompson, G., & Bennett, J. (2013). Science Teaching and Learning Activities and Students' Engagement in Science. *International Journal of Science Education*, *35*(8), 1325–1343. <https://doi.org/10.1080/09500693.2011.608093>
- Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis* (Vol. 608). Springer.
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47.

- House, J. D. (1995). The predictive relationship between academic self-concept, achievement expectancies, and grade performance in college calculus. *The Journal of Social Psychology, 135*(1), 111–112.
- Johnson, D. W., & Johnson, F. P. (1991). *Joining together: Group theory and group skills*. Prentice-Hall, Inc.
- Johnson, D. W., & Johnson, R. T. (2002). Learning Together and Alone: Overview and Meta-analysis. *Asia Pacific Journal of Education, 22*(1), 95–105.
- Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review, 19*(1), 15–29.
- Jungić, V., Kaur, H., Mulholland, J., & Xin, C. (2015). On flipping the classroom in large first year calculus courses. *International Journal of Mathematical Education in Science and Technology, 46*(4), 508–520.
- Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology, 64*(5), 402–406. <https://doi.org/10.4097/kjae.2013.64.5.402>
- Karjanto, N. (2017). Attitude toward mathematics among the students at Nazarbayev University Foundation Year Programme. *International Journal of Mathematical Education in Science and Technology, 48*(6), 849–863.
- Kogan, M., & Laursen, S. L. (2014). Assessing Long-Term Effects of Inquiry-Based Learning: A Case Study from College Mathematics. *Innovative Higher Education, 39*(3), 183–199. <https://doi.org/10.1007/s10755-013-9269-9>
- Lasserre, P. (2009). Adaptation of team-based learning on a first term programming class. *ACM Sigcse Bulletin, 41*, 186–190.
- Laursen, S. L., Hassi, M.-L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education, 45*(4), 406–418.
- León-Mantero, C., Casas-Rosal, J. C., Pedrosa-Jesús, C., & Maz-Machado, A. (2020). Measuring attitude towards mathematics using Likert scale surveys: The weighted average. *PLOS ONE, 15*(10), e0239626. <https://doi.org/10.1371/journal.pone.0239626>
- Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics, 82*(1), 145–164.
- Lindaman, B. J. (2007). *Making sense of the infinite: A study comparing the effectiveness of two approaches to teaching infinite series in calculus* [PhD Thesis]. University of Kansas.

- Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, 74(2), 118–122.
- Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, 26–47.
- Maggelakis, S., & Lutzer, C. (2007). Optimizing Student Success in Calculus. *PRIMUS*, 17(3), 284–299. <https://doi.org/10.1080/10511970601182764>
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95(5), 877–907. <https://doi.org/10.1002/sce.20441>
- McLeod, D. (1994). Research on affect and mathematics learning in the JRME: 1970 to present. *Journal for Research in Mathematics Education*, 25(6), 637–647.
- McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20.
- McNeish, D., & Stapleton, L. M. (2016). Modeling Clustered Data with Very Few Clusters. *Multivariate Behavioral Research*, 51(4), 495–518.
- Ngurah, A., & Lynch, D. P. (2013). A confirmatory factor analysis of attitudes toward mathematics inventory (ATMI). *The Mathematics Educator*, 15(1), 121–135.
- Niederle, M., & Vesterlund, L. (2011). Gender and Competition. *Annual Review of Economics*, 3(1), 601–630.
- Otero, V., Pollock, S., & Finkelstein, N. (2010). A physics department's role in preparing physics teachers: The Colorado learning assistant model. *American Journal of Physics*, 78(11), 1218–1224.
- Paulson, D. R. (1999). Active learning and cooperative learning in the organic chemistry lecture class. *Journal of Chemical Education*, 76(8), 1136.
- Pek, J., Wong, O., & Wong, A. C. M. (2018). How to Address Non-normality: A Taxonomy of Approaches, Reviewed, and Illustrated. *Frontiers in Psychology*, 9, 2104.
- Pell, G. (2005). *Use and misuse of Likert scales*.
- Popham, W. J. (2005). Students' attitudes count. *Educational Leadership*, 62(5), 84.
- Primi, C., Bacherini, A., Beccari, C., & Donati, M. A. (2020). Assessing math attitude through the Attitude Toward Mathematics Inventory – Short form in introductory statistics course students. *Studies in Educational Evaluation*, 64, 100838.

- Pyzdrowski, L. J., Sun, Y., Curtis, R., Miller, D., Winn, G., & Hensel, R. A. (2013). Readiness and attitudes as indicators for success in college calculus. *International Journal of Science and Mathematics Education*, 11(3), 529–554.
- Rasmussen, C., Marrongelle, K., & Borba, M. C. (2014). Research on calculus: What do we know and where do we need to go? *ZDM*, 46(4), 507–515.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>
- Reyes, L. H. (1984). Affective Variables and Mathematics Education. *The Elementary School Journal*, 84(5), 558–581.
- Rubin, D. B. (1975). Inference and missing data. *ETS Research Bulletin Series*, 1975(1), i–19.
- Sadler, P., & Sonnert, G. (2018). The path to college calculus: The impact of high school mathematics coursework. *Journal for Research in Mathematics Education*, 49(3), 292–329.
- Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review*, 19, 158–172. <https://doi.org/10.1016/j.edurev.2016.07.003>
- Saxe, K., Braddy, L., Bailer, J., Farinelli, R., Holm, T., Mesa, V., Treisman, U., & Turner, P. (2015). *A common vision for undergraduate mathematical sciences programs in 2025*. Mathematical Association of America.
- Slavin, R. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, 21, 43–69.
- Sonnert, G., Barnett, M. D., & Sadler, P. M. (2020). The effects of mathematics preparation and mathematics attitudes on college calculus performance. *Journal for Research in Mathematics Education*, 51(1), 105–125.
- Sonnert, G., Sadler, P. M., Sadler, S. M., & Bressoud, D. M. (2015). The impact of instructor pedagogy on college calculus students' attitude toward mathematics. *International Journal of Mathematical Education in Science and Technology*, 46(3), 370–387.
- Sorge, C., & Schau, C. (2002). Impact of engineering students' attitudes on achievement in statistics: A structural model. *Annual Meeting of the American Educational Research Association*. New Orleans.

- Stanic, G. M., & Hart, L. E. (1995). Attitudes, persistence, and mathematics achievement: Qualifying race and sex differences. *New Directions for Equity in Mathematics Education*, 2, 258–276.
- Streiner, D. L. (2015). Best (but oft-forgotten) practices: The multiple problems of multiplicity—whether and how to correct for many statistical tests. *The American Journal of Clinical Nutrition*, 102(4), 721–728.
- Tapia, M. (1996). *The Attitudes toward Mathematics Instrument*. [Paper presentation]. 1996 Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, AL. <https://eric.ed.gov/?id=ED404165>
- Tapia, M., & Marsh, G. E. (2002). *Confirmatory Factor Analysis of the Attitudes toward Mathematics Inventory*. Chattanooga: Paper presented at the Annual Meeting of the Mid-South Educational Research Association.
- Tapia, M., & Marsh, G. E. (2004). An instrument to measure affect. *Mathematics Education Quarterly*, 8(2), 56–62.
- Thiele C, Hirschfeld G (2021). cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *Journal of Statistical Software*, 98(11), 1–27.
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399.
- Yee, L. S. (2010). Mathematics Attitudes and Achievement of Junior College Students in Singapore. In *Mathematics Education Research Group of Australasia*. Mathematics Education Research Group of Australasia.
- Yin, J., & Tian, L. (2014). Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. *Computational Statistics and Data Analysis*, 77, 1–13.
- Zan, R., Brown, L., Evans, J., & Hannula, M. S. (2006). Affect in Mathematics Education: An Introduction. *Educational Studies in Mathematics*, 63(2), 113–121.