THE IMPACT OF INCREASING MIDDLE SCHOOL SCIENCE CLASS SIZE ON STUDENT ACHIEVEMENT AND TEACHER BURNOUT

by

Rachael Nicole Gradwell

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2025

DEDICATION

To my mom, who believes in me endlessly.
To my husband, who supported me through it all.
To my son, my sunshine.
You gave me the heart needed to accomplish this.
Thank you.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my husband, Drew Gradwell, for his unwavering support, thoughtful feedback, and keen eye in proofreading my work. Thank you for all the weekend moments where you carved space for me to complete my work, juggling all the needs and allowing me to be the passenger princess typing away on road trips.

I am also grateful to Josh Dennis for his nod to the program and for his valuable capstone feedback. Nathan Blass for his expert editing guidance and Amy Shearer for her insightful numbers feedback.

I extend the gratitude to my science reader, Dave Willey, for his thoughtful feedback which strengthened my research, and to Walt Woolbaugh for his immense guidance and abundant feedback throughout this process.

I also appreciate all my fellow peers in the MSSE discussion boards for the enlivening conversations that propelled me forward.

Finally, I want to acknowledge my students, who inspire me daily and make many of my days full of joy.

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

One problem currently plaguing many school districts is the shortage of qualified teaching personnel. As a partial solution, some districts have turned to increasing class sizes. This study sought to explore the impact of increasing middle school science class size on student achievement, behavior, work completion, and teacher burnout. During the study, the teacher's student load was increased by more than 50% from 79 to 124 students. During this time, a planned student-teacher practicum occurred in the classroom as well. A mixed-methods approach was implemented and gathered data included gradebook analysis, behavioral referral tracking, student surveys, burnout scales, student interviews, and teacher journals. Quantitative data were analyzed for standard deviations, averages, medians, and quartile differences. Qualitative data were coded for recurring themes and were used to help triangulate claims. The increase in class size impacted some students more than others. The top 25% showed little to no change in academic performance with the larger class sizes. The lower 75% of learners registered more differences in academic impact due to the change. Student data revealed decreased on-time work completion and teacher burnout soared in the larger group. The support of the student teacher had a modest effect on work completion and lessened burnout somewhat. Overall, the study concluded that increased class sizes do not affect all students equally and affirmed that larger class sizes contribute

CHAPTER ONE

INTRODUCTION AND BACKGROUND

It was a year like no other in my school. On every team we had missing teachers and had classrooms that sat empty and unused awaiting a spark of inspired instruction and a teacher who cared. We were not alone in this in a nation currently struggling with a teacher shortage. Many districts choose, as mine did, to combine classes and make larger groups under one educator. The purpose of this AR project was to examine the possible ramifications of that decision. It also aimed to better understand how to mitigate the impact on students and teachers by identifying how it affects them and who it affects most in a middle school classroom setting. The central question guiding this investigation was: *How do variations in classroom size and the presence of a student teacher affect student's achievement, work completion, and behavior?* The question has also been driven deeper with the following sub questions:

1. What impact did the increased class size have on feelings of teacher efficacy and burnout?
2. When both educators were present, student teacher and original teacher, did the extra support mitigate the impact of the increased class size or did any disparities remain?
3. What impact did the changes in class parameters have on student perspectives of teacher availability and support?

The researcher was placed into a unique space to study this when an abrupt shift occurred mid-year with a month's notice. The total student load within the teacher's roster increased by 57% going from 79 students to 124. The class parameters like number of assignments and instruction remained the same and, thus, a unique situation to study this change on the original

students was created. The treatment was broken into four phases following the students on their journey of classroom shifts. The first phase created a solid baseline for comparison as it consisted of the original teacher and the small group of students. The second phase introduced a student teacher to the mix with the small group. The third phase included the transition to the larger group while the student teacher was present. This phase gave us insights into whether the additional support could effectively counterbalance the impact of larger class sizes. In the final phase, the original teacher managed the increased load of students alone. This study can be classified as a mensurate experiment.

To examine these factors in depth, a variety of data collection tools were developed. As part of a true action research project, the collection tools were both qualitative and quantitative in nature. Quantitative measures included a comprehensive gradebook analysis, work completion tracking, the Zager Teacher Burnout Scale, and post-testing growth comparisons. Qualitative measures included teacher journaling and student interviews which both provided important depth and nuance to this research.

The researcher's team consisted of three co-workers and a professor at Montana State University. All members of the team have given valuable feedback when asked. The researcher has asked members questions on formatting, designing collection tools, and analyzing the data. Although roles were originally designated, the team's inputs were more fluid than originally anticipated and were based on the workload and availability of the individuals. Each helped as they could.

CHAPTER TWO

CONCEPTUAL FRAMEWORK

Overview of Findings on Class Size Reduction

Policy makers, teachers, researchers, and school boards have long debated the relationship between class size and student achievement and many studies revealed mixed outcomes. In 2018, a systematic review of 148 reports from 41 countries was conducted by the Campbell Collaboration Group. After analyzing the reports, they concluded that smaller class sizes do make some positive difference in reading achievement in younger students and especially for students who experience adversity in other areas. However, it was noted in the review that the "effect on mathematics achievement was negative and not statistically significant" (Filges, 2018, p. 51). The effects also appear to diminish with student age. Because the overall impact was modest and somewhat inconsistent, other factors like teacher quality may be more important. Even among the reviews, a major trend found when studying the potential impact of class sizes seems to be the inconsistency of the results on all students.

The Campbell study included 45 studies that referenced Tennessee's STAR (Student Teacher Achievement Ratio) experiment from the 1980s. This is a foundational study on class size reduction (CSR). The study included real reductions of class size taking classes from an average of 24 pupils down to 15 and included 6000 students followed over a four-year time period- from kindergarten through grade four. The study "yielded an array of benefits of small classes including improved teaching conditions, improved student performance during and after the experimental years, improved student learning behaviors, fewer classroom disruptions and

discipline problems, and fewer student retentions" (Finn, 1999, p. 98). Because of the plethora of positive effects, many follow-up studies were done to try to identify the exact change that simply the number of students was having on the classroom characteristics. They found that teachers do not change their teaching strategies with smaller groups, and no statistical differences were found in most teacher activities or classroom atmosphere. The most profound changes appear to be in student behavior and engagement. Students in smaller classes had more "contacts with the teacher for purposes of classification, giving answers to questions that were open to the whole class, and contacting the teacher privately for help" (Finn, 1999, p. 103). Students in smaller groups were more on task and more connected to teaching and learning interactions. Concepts from social psychology that could explain this connection include the "social loafing" and "diffusion of responsibility" that occurs in larger groups (Levine, 1998, Chapter 16).

Hanushek argues the opposite of the STAR study stating that the aggregate data shows that "pupil-teacher ratios and class sizes have fallen dramatically over the past three decades and student performance has remained virtually unchanged." He examined many other demographic changes in the country that may have contributed to a net neutral effect of the smaller sizes in the modern era such as lessened support from home, the changing structure of schools, and the differences from country to country where similar class sizes are found. He concludes that the cost of CSR is great and, when examining the large-scale picture, there "appears to be little systemic gain from general reductions in class size" (Hanushek, 1999, p. 160). These findings of opposing conclusions further give reason to study this within the educator's current setting and demographics. This is particularly important because, as summarized from Chingos (2013), the cost of reducing class sizes is unlikely to be as beneficial as putting that funding towards other

proven programs like “cross-age tutoring, early childhood programs, and increases in instructional time” (para. 84). Research in our schools helps make decisions for our schools.

Class Size and Teacher Perspectives

From the report *Rethinking Class Size* by Blatchford and Russell (2020), teachers of large class sizes (upwards of 30 students) reported adverse consequences including “feelings of guilt, stress, tiredness, less creative energy, and poor health.” This is of great concern for our students as a multinational study demonstrated that teacher “morale and motivation generally have a significant role on students science and math achievement in all countries” (Abazaoglu, 2016, p. 2616). That study also concluded that a teacher’s motivation was a much stronger factor in areas where there is a greater disparity in income among classes. This finding underscores the need for policies and interventions that focus on supporting teachers, particularly in high poverty schools where the higher class sizes, lack of resources, and other challenges may have the highest impact on their motivation (Lochmiller, 2024).

Teacher Burnout is a trending term for a reason. In a report on teacher well being published in January 2024, it was reported that “compared with comparable working adults, about twice as many teachers reported experiencing frequent job-related stress or burnout and roughly three times as many teachers reported difficulty coping with job-related stress” (Doan et al., 2024, para. 3). A study done by Saloviita (2021) found that class size did have some impact on burnout. Classroom teachers with a higher student load reported higher rates of exhaustion than teachers with a lower load. The difference Saloviita (2021) found was not as pronounced as many had expected. Work published in the IOER International Multidisciplinary Research Journal further uncovered significant differences on the impact of class size on teachers’

emotional health based on gender, years of experience, age, and level of educational attainment. Some factors, like educational attainment, were protective against burnout, while others, like age, exacerbated the negative effect of the change in class size (Manlongat et al., 2021). This suggests that the impact on teachers can be magnified on some individuals while being minimal on others and further gives reason for this study.

Student Perspectives of Teacher Availability and Support

Among our goals as educators is not simply for students to learn standards, but also for them to enjoy school and the educational process. Studies on student enjoyment found factors such as "successful social relationships, small classes, variation in learning" and teacher care and support to be highly correlated with higher perspectives and engagement with the educational process (Gorard, 2011, p. 681). In other studies, it has been shown that as class sizes get relatively larger, students felt they received less teacher interaction and support and reported lower levels of satisfaction with the class and their learning (Wang, 2022).

CHAPTER THREE

METHODOLOGY

Treatment

Variations to the class size occurred due to district needs, as did the placement of the student teacher. In this scenario, the researcher utilized significant change to the student roster size as the impetus for studying said change. This variation was the central piece of the treatment. The treatment window was lengthened and approved to use data generated before the change occurred. Thus, the comparison group was the original small group of students prior to the student teacher joining the class. Next, the student teacher joined, and a second cohort of data was generated. In the third phase of the treatment, the student number was increased by 57%. In this phase, the number of students under this educator went from 79 to 124. In the final phase of the treatment, the number of students stayed at 124, but the student teacher went on to the next stage of her education. The difficulty of the course, requirements for assignments, and general rigor was maintained throughout all phases of the treatment.

The researcher completed the IRB Approval process in spring of 2024 (Appendix A)

Table 1. Variations in treatment paired to interval lengths.

Interval Length: 23-24 School Year	Variation in Class Treatment
September 4- January 22	79 students (over four class periods) with one classroom teacher
January 23- February 26	79 students with a student teacher and a classroom teacher circulating/helping/observing
February 26 - April 8	124 students with a student teacher and a classroom teacher circulating/helping/observing
April 12- May 9th	124 students with one classroom teacher

Demographics of Sampled Population

The demographics of the sampled population are generally reflective of Dover Area School District as a whole and represented a fair and diverse sample. Parent support can be difficult to sustain in our community and larger numbers of permission slips were unable to be obtained and returned. Three attempts were made per student with simple emails and phone calls attempted as well. This number represents the students who were able to be fully permissioned for the research.

Table 2. Sample sizes by class period.

Class Period	Number of Sampled Students
Period 1	14
Period 3	5 (smallest original class)
Period 4	17
Period 7	10

A total of 46 students were enrolled into the study. This represents about 58% of the original population who experienced the increased class sizes and thus allowed, in the researcher's opinion, a valid sample to work with during analyses.

Table 3. Ratio of learning support students to general education.

Number of Learning Support Students	General Education Students
10	36

According to 2024 State of Education Report, 19.3% of the student population receives special education programs and services (Pennsylvania School Boards Association [PSBA], 2024, p. 12). Dover's demographic of students receiving services has always been somewhat higher than the overall state sample, which was discussed at a recent faculty meeting. For my study, the ratio of permissioned students (show above, Table 3) closely reflects my school's population of 21% learning support students.

Table 4. Race and gender demographics.

Race	Male	Female	Non-Binary/Other	Total
White/Caucasian	14	18	1	33
African-American/Black	3	2	–	5
Hispanic/Latino	1	3	–	4
Other/Multiracial	1	3	–	4
	Total: 19	Total:26	Total:1	46

Dover Area School District is a typical, rural Pennsylvania school. The student population is 75.6% white and has a 24.4% enrollment in minority populations (U.S. News &

World Report, n.d.). The study sample closely aligns with the district with 28% of students identifying as multiracial, Hispanic/Latino, and Black.

Data Collection and Analysis Strategies

Three major variables were targeted for data collection to ensure fidelity to the core AR question: student achievement, work completion, and student behavior. Data were compared for each across four different periods: small group no student teacher, small group student teacher, large group student teacher, and large group no student teacher.

For student achievement, the primary measure was an analysis of changes within students' scores across phases of treatment in the class gradebook. A duplicate of the class gradebook for the entire year was made in Google Sheets (Google.org) on the secure school server with students de-identified. Only permissioned students were included. The researcher calculated descriptive statistics including averages, medians, ranges, and standard deviations for each assignment and each period and then compared those from one period to the other in the search for significant differences. This analysis provided a solid source of evidence on student achievement within class assignments. Susan Brookhart's 2015 synthesis of 28 studies on using grades to examine achievement noted that class achievement could differ from tested achievement; therefore, a comparison was also made between historical post-testing growth per unit and the treatment year's post-testing growth per unit to analyze whether students' achievement of standards differed significantly. Two independent t-tests were run on the tested achievement results. One included the final scores of both groups. The second included the growth between the pre-test and the post-test for the historical group and the treatment group.

For work completion, a three-column data table was created for all assignments and separated by phase of treatment. The three columns included number of assignments turned in on time, number turned in late, and number never completed. Differences from one period to another in percentage per category helped illustrate the impact on student work completion with more or less help available if a difference exists. Work completion data also illustrated important information about time on task and engagement with content as opposed to off-task behavior (Herman, 2020).

For student behavior, examining the number and type of office referrals is a common practice to monitor changes in behavior. Previous literature has found referral reports and patterns within them to be a "valid indicator of student behavior and school functioning" (Lassen, 2006, p. 703). For this study, an anonymized referral report has been sent to the researcher from the school guidance counselor that listed all the referrals generated by the researcher in response to student behavior during the treatment period. The report included both date and type of referral. An in-depth look at the report revealed a changed pattern of behaviors, by frequency or type, from one interval to another.

The subquestions generated by the core AR need additional data to be fully addressed. When examining teacher feelings of motivation and efficacy, the 21 question Teacher Burnout Scale created by Zager and Seidman is a crucial instrument (Appendix B). It was administered at three intervals: three weeks into the transition when the student teacher was still present, three weeks after the student teacher left, and three weeks into the new school year with smaller classes. Teacher journaling also provided valuable insight into this sub question. The journaling occurred on a bi-weekly frequency with a template format to facilitate further analysis. The

template consisted of three questions related to the target subquestion: How well have I met student needs this week, What successes and challenges have I encountered, and How do I feel about the week of teaching ahead?

The possible mitigating effect generated by the student teacher would appear as a difference between the large group with the student teacher and the large group with only one teacher. Both qualitative and quantitative data has been compared to see if there is any clear alignment between the student teacher's presence and severity of impact.

Students' perspectives were examined using interviews post-treatment. The interview consisted of three main questions with probing sub questions (Appendix C). These questions addressed changes in class size and teacher support, opportunities for asking questions and getting help, preferences for class size, and the perceived impact of these factors on student learning and comfort. The interviews were coded thematically to show recurring patterns. Quantitative calculations were applied where applicable including determining the percentage of students who preferred small groups and reported impacts on focus. Illustrative quotes were pulled as evidence as well. Analysis of the interviews followed rigorous guidelines based on field norms to help prevent seeping of the researcher's own views into the results and should provide solid qualitative and quantitative data. (Braun, 2006)

Validity was ensured throughout the study in several ways. For each of the questions examined by the researcher, qualitative and quantitative instruments were aligned to allow for triangulation. Instruments were chosen carefully based on peer and professor feedback. When possible, vetted instruments like the Seidman and Zager Teacher Burnout Scale were utilized.

Table 5. Research matrix.

Area	Data Collection and Research Methods						
	Comparative Gradebook Analysis	Post Testing- Growth compared to previous years	LMS- on time, late, and never submitted data	Anonymized report through central office	Seidman and Zager Teacher Burnout Scale - 3 administrations	Teacher Journaling	Post Course Student Interviews
Core Question: How do variations in classroom size and the presence of a student teacher affect student achievement, work completion, and behavior in a middle school classroom?							
	-Focus: Student Achievement						
	X	X					X
	-Focus: Work Completion						
			X				X
	-Focus: Student Behavior						
				X		X	X
Sub Questions							
What impact did the increased class size have on feelings of teacher efficacy and motivation?							
					X	X	
When both educators were present, student teacher and original teacher, did the extra support mitigate the impact of the increased class size or did any disparities remain?							
	X		X	X	X	X	
What impact did the changes in class parameters have on student perspectives of teacher availability and support?							
						X	X

CHAPTER FOUR

DATA ANALYSIS

Results Regarding Student Achievement

The students' percentage scores retrieved from the gradebook provided a good foundation for exploring trends in class grade achievement across the four phases of treatment. Surprisingly, the students scored the highest in the large group with one teacher phase with a mean of 88%, a median of 93.3%, and the lowest variability (standard deviation 0.15). Taken broadly, these metrics initially suggest that students in this phase had relatively high and consistent achievement. Students scored the lowest in the large group with two teachers phase recording both their lowest mean (81.8%) and lowest median score (85%).

Table 6. Whole group descriptive statistics based on class assignment grades, (*n*=46).

Phase of Treatment:	Small group with one teacher	Small group with a student teacher and the original teacher	Large group with student teacher and the original teacher	Large group with one teacher
Mean	86%	83.93%	81.8%	88%
Median	90.00%	90.00%	85.00%	93.33%
Standard Deviation	0.17	0.18	0.17	0.15

Breaking the data into quartiles revealed important insights. In all methods of analysis, small group with one teacher and large group with one teacher are close mirrors. Even when broken into the lowest quartile, middle quartiles, and highest quartile we see the most similarities in those two phases of treatment. For the top 25% of students, we see performance stay

consistently high across all phases of treatment. There is a slight dip in the initial transition to a large group, but they rebound well in the last phase. For the middle 50%, performance drops significantly in the large group with one teacher phase, and they show more variability than the top performers across stages. The lowest 25% showed the most variability across stages. They drop when the student teacher initially begins to take over and again when the large group with both teachers occurs with their mean weighing in at a 69% average during that phase. Students interviewed from that quartile refer to that phase of class time as feeling "less focused", "busier", and "crowded, because, ya know, there were kids everywhere."

Table 7. Medians across quartiles, (*n*=46).

Medians Across Groups	Small group with one teacher	Small group with a student teacher and the original teacher	Large group with student teacher and the original teacher	Large group with one teacher
Top 25%	100.00%	99.00%	95.50%	100.00%
Middle 50%	90.00%	90.00%	85.00%	93.33%
Bottom 25%	80.00%	80.00%	75.50%	80.00%

Analysis of medians further supports that the performance patterns for small group one teacher and large group with one teacher mirror one another. Across those two stages, only a 3% difference in the middle quartiles' median was observed.

Table 8. Standard deviation across quartiles, (*n*=46).

Standard Deviation Across Groups	Small group with one teacher	Small group with a student teacher and the original teacher	Large group with student teacher and the original teacher	Large group with one teacher
Top 25%	0.09	0.08	0.09	0.06
Middle 50%	0.15	0.12	0.14	0.12
Bottom 25%	0.20	0.26	0.22	0.20

The standard deviation increased for each quartile doubling or, in some cases, tripling from the top quartile to the lowest quartile, showing that top students were more analogous to one another than bottom students were. This further demonstrates the stability of the top group and the magnifying effect of the treatment on individuals within lower groups. For some individuals in the bottom 25% of students, the transitions to the student teacher and the large group with the student teacher were very difficult, and significant drops in scores were observed.

The box plots created help illustrate the distribution of assignment scores for the Top 25%, Middle 50%, and Bottom 25% across the four phases of treatment. For the top 25%, the median remains consistently high with very little fluctuation between phases. The group also showed very narrow interquartile ranges, indicating minimal variability in performance as shown previously through the low standard deviation and consistent medians. Patterns analyzed in the interviews showed that this group was more likely to view the transition to larger classes as positive citing the diversity of opinions and the perceived increased opportunities to collaborate as features that helped them understand concepts better. One student in the high-achieving group summarized this idea stating that “There was more people to talk to in labs and groups -more

people to talk with and opinions to be heard. If there's more opinions, you understand the topic more deeply with more viewpoints."

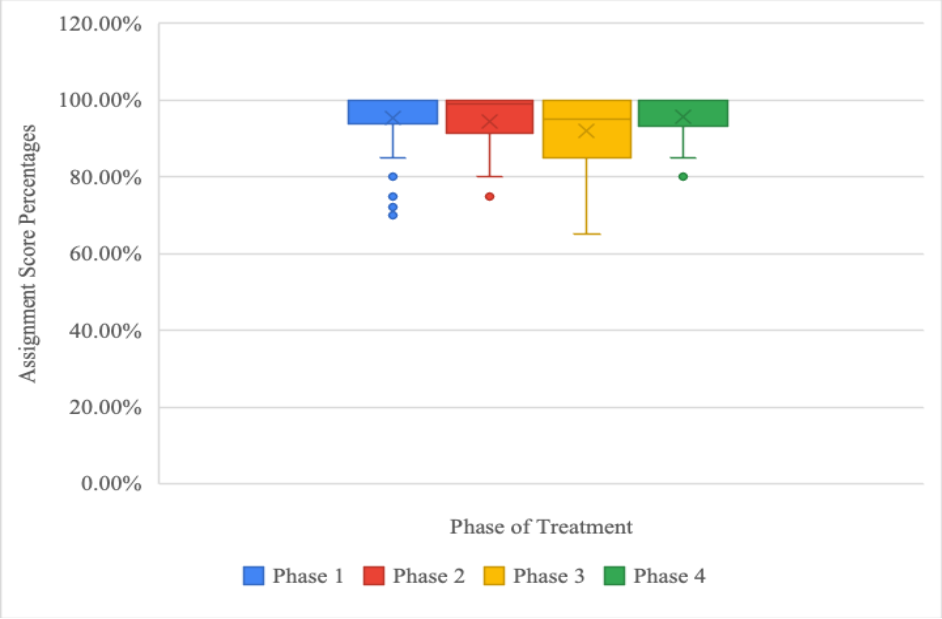

Figure 1. Top 25% of students' assignment score distribution across treatment, (*n*=11).

The Middle 50% demonstrated more noticeable shifts. The median remained stable at 90% in the small group phases but dropped to 85% in the large group during the two teachers phase. There were more outliers in this phase, indicating higher variability in performance across the group than the top learners (Fig. 2).

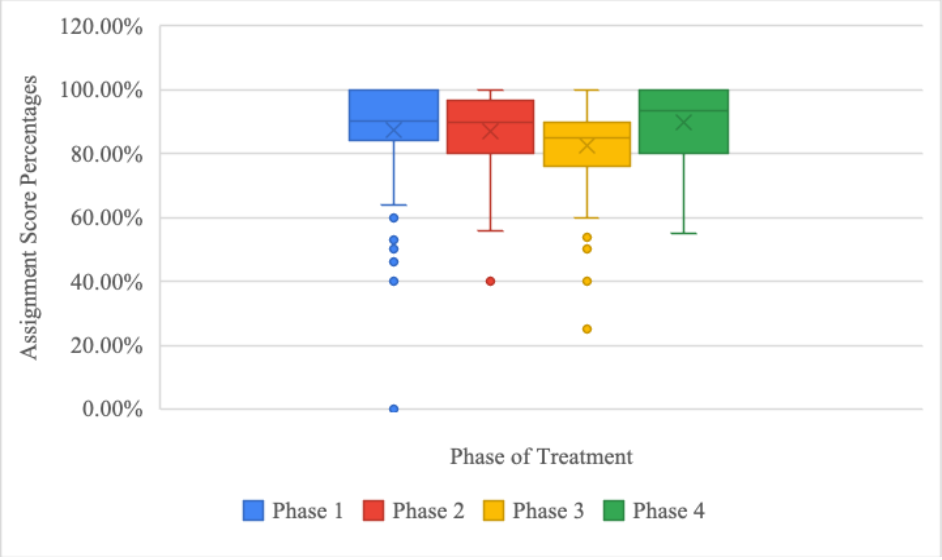

Figure 2. Middle 50% of students' assignment score distribution across treatment, (*n*=24).

The bottom 25% exhibited the most variability. In the small group phase with both teachers and the large group with both teachers, the interquartile ranges were wide, and the whiskers extended significantly, showing that student performance was inconsistent and some students struggled greatly. In interviews, students highlighted challenges with accessing teacher support in the larger group noting that longer wait times delayed their ability to resolve concepts they were stuck on. One student reflected, "You [the teacher] tried helping everyone whenever you could, but wait times were longer because there were more people," while another added, "I was raising my hand for longer, and no one came."

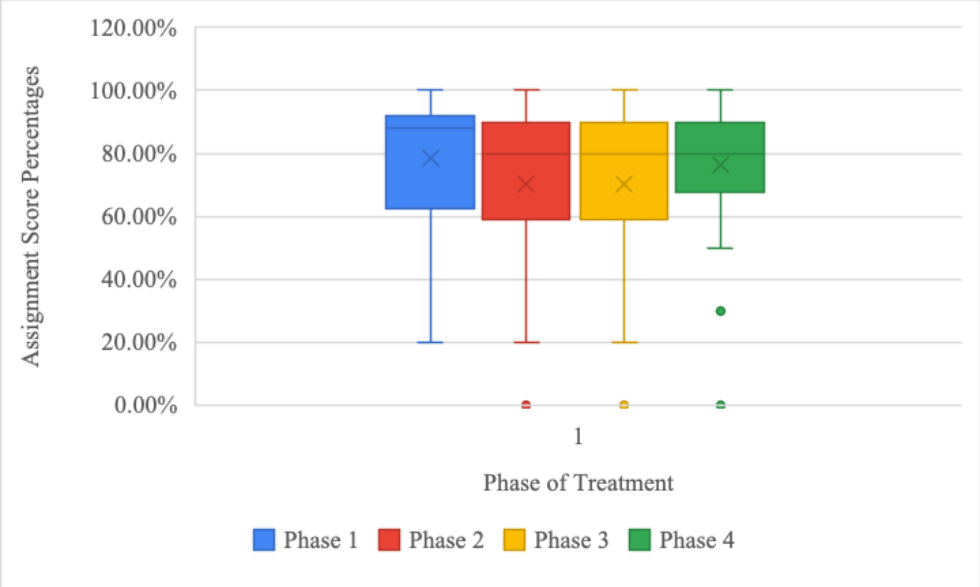

Figure 3. Bottom 25% of students' assignment score distribution across treatment, (*n*=11).

Comparison of standardized course post-testing results from the treatment year to historical years offered a broader perspective on achievement and growth. An independent t-test comparison across both years yielded a high p-value (0.488), indicating no statistically significant difference in standard achievement from the treatment. A comparison on the percent of growth students achieved came closer to significance with a p-value of 0.137, but the small effect size of that growth (Cohen's d= 0.276) demonstrates that the growth differences were not robust enough or widespread enough to achieve statistical significance. The stable performance of the top learners may have masked differences once the data was taken at the aggregate level. Additionally, the steps taken to prepare students for the standardized testing such as a full course review designed to fill in gaps, may have skewed the data set.

Results Regarding Work Completion

Analysis of aggregate, anonymous data gleaned from the learning management system revealed a striking trend in timely work completion rates across the four phases of treatment. Unlike the initial data regarding student achievement, a clear difference in work completion rates was immediately evident as group size and teacher support changed.

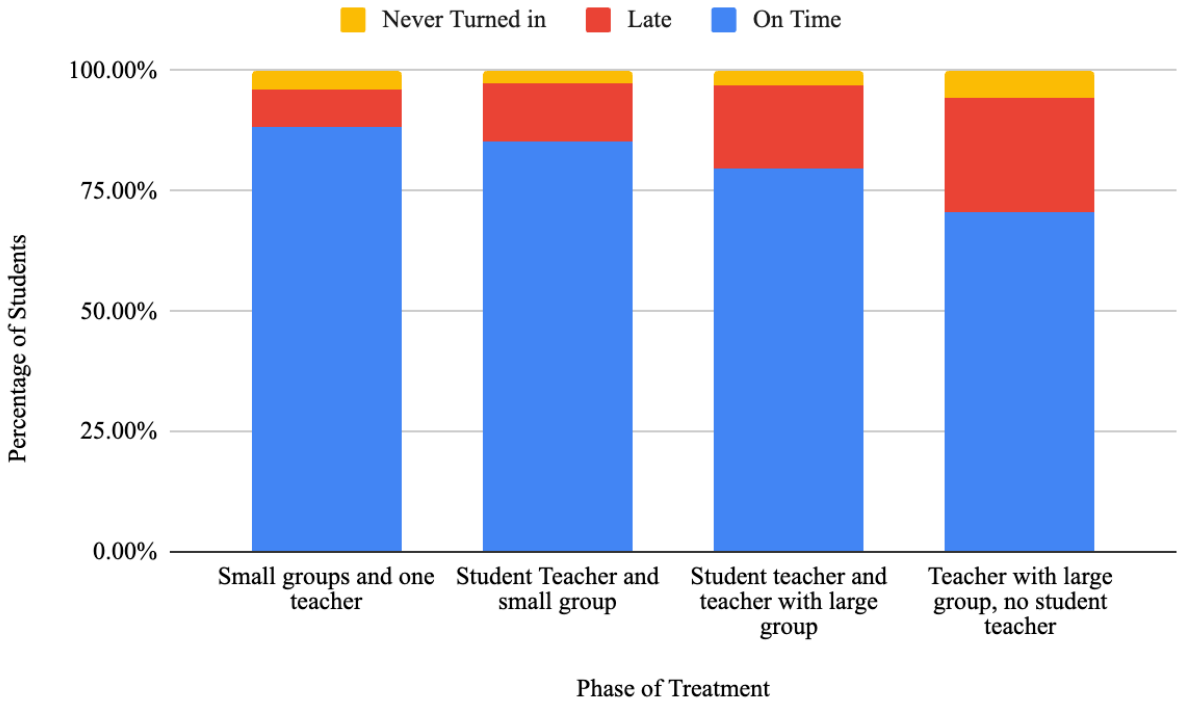

Figure 4. Timeliness of work across treatment phases, (*N*=124)

On-time work represents work that was finished by its due date and submitted in the appropriate way. As the group sizes increased and additional teacher support was reduced, the percentage of on-time submissions decreased. Students achieved the highest rate of on-time

submissions at 88.49% during the small groups with one teacher phase. This figure represents a stark contrast to the 70.59% on-time submissions rate observed during the large groups with one teacher phase. In the largest group, students were significantly less likely to submit their work punctually. In regard to work completion, having the student teacher present in the room did seem to slightly mitigate this decline as the on-time submission rate during that phase was 79.50%, suggesting that the additional support did help students in the larger setting.

In the Dover Area School District, school policy allows for assignments to be turned in late for up to two weeks past the assignment date. Late assignments rose in coordination with the number of students. In the small groups with one teacher phase, only 7.67% of assignments were late. Late assignments were generally rare for the average student. During the large group with one teacher phase, the number of late assignments rose sharply to 23.7% late, marking nearly 1 in every 4 assignments as being turned in late. The disconnect between the rise in late work completion rates and the relatively neutral shift in overall academic achievement across the same two phases is likely due to the minimal late penalty applied to student work, as well as differences in performance across quartiles. Student interviews show a pattern of lessened support being recognized by students as the group size increased. This lessened support was the likely catalyst for the increase in the late submissions. One student reported "there was less help because there were more kids to get to for questions. I realized I had to depend on myself more for figuring things out."

The “never turned in” assignment category represents assignments that were never completed by students or turned in for any credit. Unsubmitted work becomes a zero in the class gradebook and drastically alters the trajectory of students’ grades for the marking period.

Although the unsubmitted work rate never went above 6%, a significant difference was shown during the treatment. In this scenario, the smallest rate of unsubmitted work- 2.61% - occurred when both the student teacher and teacher were present with the small group. That rate more than doubled to 5.71% during the large group with one teacher phase. Students expressed difficulty about asking for help when struggling with the work in the larger group, as shown in this quote from a student interview: "I barely asked questions after the switch because I was shy and had social anxiety. I was fine before the combination, but the crowding changed my feeling."

Results Regarding Student Behavior

As class size increased, a complex shift in student behavior and focus was observed. Analysis of student interviews revealed that 55.6% of students felt that their focus decreased with the larger class, 27.8% reported no change, and 16.7% felt that their focus improved. Students who reported a positive change in focus were more likely to view their peers as supports to them in accessing the content. One student assessed the situation by saying that “it was more fun to have more people in class. There were more people to talk and work with and some of my friends got switched to that class." The students who felt their focus decreased were likely to call the classroom noisier and to view the additional peers as distractions. Both teacher journals from the phase and the dropping work completion rates suggest that the level of conversations happening during the larger class work time was not conducive to effective time management. Results received from an anonymous report generated by our school guidance counselors echoed the importance of adult support for managing adolescent behavior.

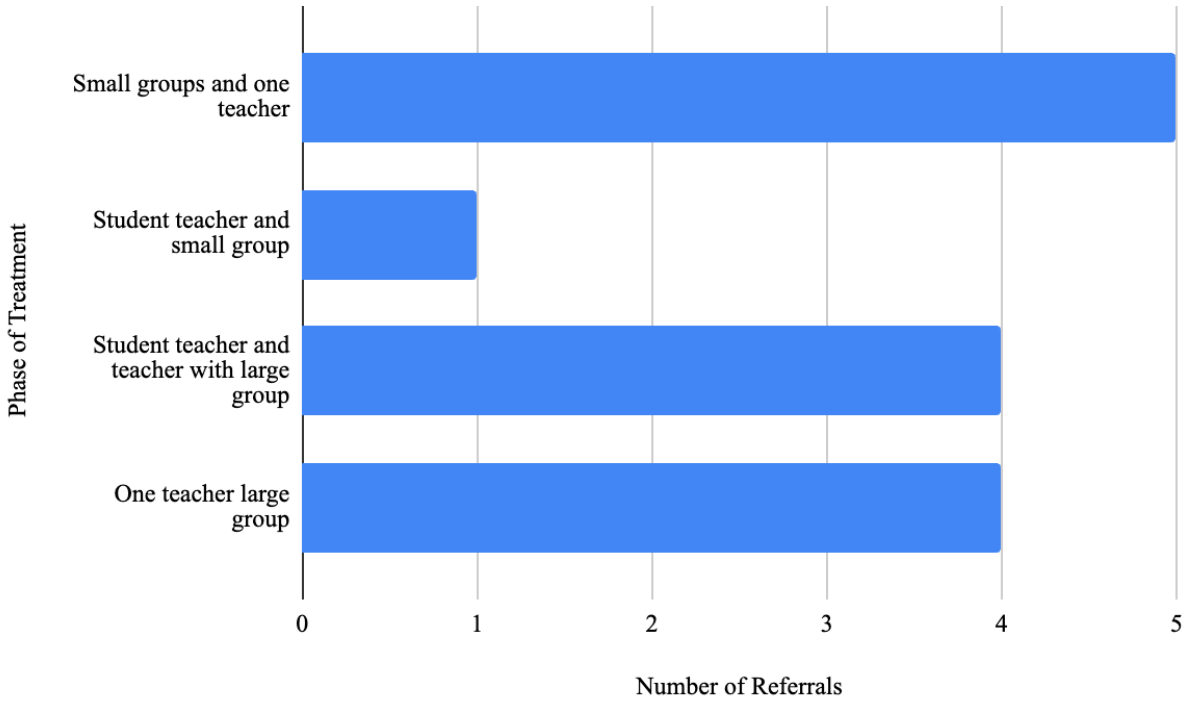

Figure 5. Student referrals across all phases of treatment, (N=124)

Referrals were rare in all scenarios, but the number of referrals and infractions were consistent across all phases but one. Only once we had two adults in the room redirecting and helping did we see a significant decline in referrals and infractions.

Results Regarding Teacher Burnout

To gather data on teacher burnout, the Seidman and Zager Teacher Burnout Scale was taken three times over the course of the study. It was first administered when the student teacher and the core teacher were working with the large group of students. The second administration was when the core teacher was managing the large group solo. The third administration occurred post-treatment when the core teacher again had a normal student load.

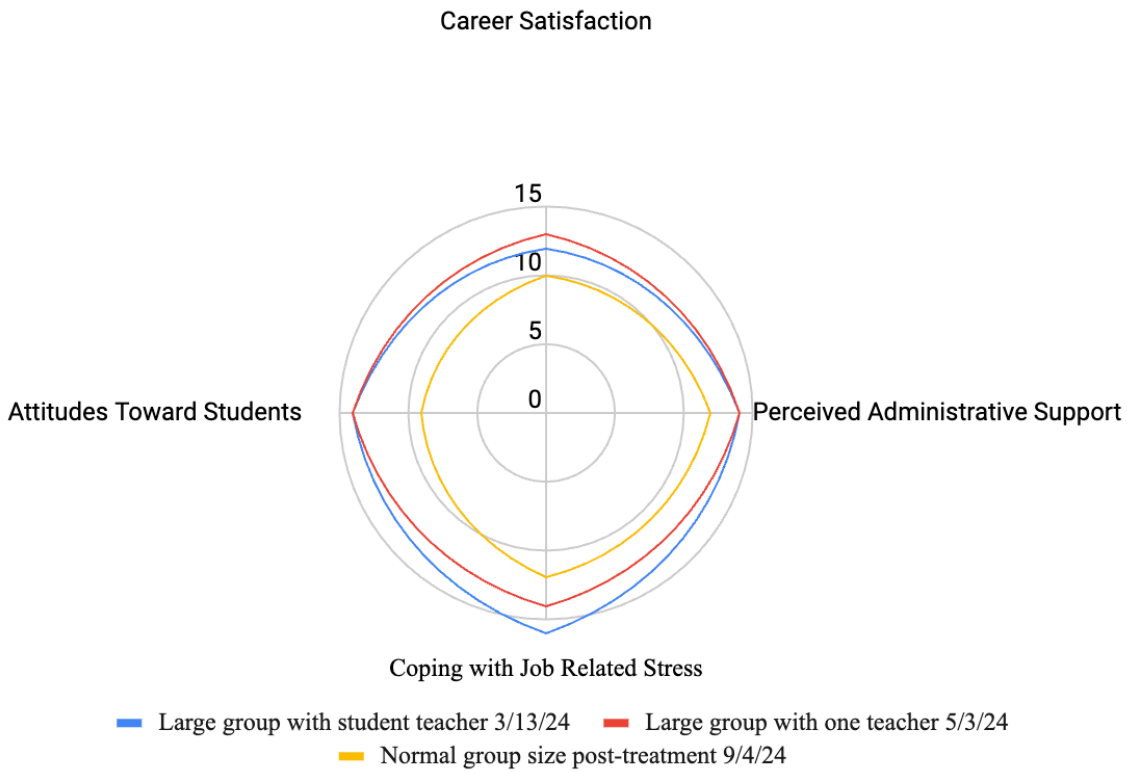

Figure 6. Teacher burnout scale results.

Shifts over the three completed scales reveal the impact of support and class size on the core teacher. For both the large group supported and large group unsupported, attitudes towards students and perceived administrative support remained steady at 14. The highest burnout score was recorded in coping with job related stress when the large group with the student teacher was present. Teacher journal entries from that time demonstrate an unease with managing many changes at once and a sense of discomfort at the pending loss of support. The teacher stated, “I find myself planning more around our departure plan for the student teacher, which has been

challenging emotionally and logistically. I'm doing my best to prepare, but I know it won't feel the same once she's gone." The burnout score topped the chart at 16 in that category at that time. So, although, the presence of the student teacher seemed to offer a benefit with career satisfaction during the transitionary period, the additional support alone did not mitigate the stress of the student increase as the support was not able to be anticipated long term.

The most significant shifts occurred when the student load decreased to a normal amount (*N*=75). By September, all signs of burnout had declined significantly. Attitudes towards students dropped from 14 to 9, representing a 35.7% reduction in burnout at that time. Job related stress decreased by 25% from the initial reception of the larger classes to the normal class sizes in the following school year.

Teacher journals provide further context to the data gathered using the burnout tools, including providing insight into the teacher's responses and causes of stress. Early entries often mentioned struggling to meet many perceived needs with one journal noting, "I wish I could be in three places at once." By the end of April, the number of entries using words like exhausted and drained peaked, and the teacher reported, "I'm doing my best, but everyday feels like a marathon." These reflections align with the quantitative data, reinforcing how increased class size contributed to stress and exhaustion—stress that significantly eased when the student load returned to a manageable level.

Results Regarding Student Teacher Influence

The student teacher already had significant classroom training at the point of the combination of both classes and was expected to be a positive asset; however, her inclusion in the transitionary phase showed mixed results. Gradebook data suggested that the additional support did not lead to higher grades for students. In fact, the lowest mean scores for all groups were recorded during the time in which the student teacher was present. Having both teachers present did help with on-time submission rates, which were significantly higher when the student teacher was present (79.5% on-time submissions in large group with two teachers) than when there was only one teacher with the large group (70.59%). This data combined suggests that the increased level of support for all students allowed for more work to get turned in, but that an additional adult without targeted intervention and with less experience did not have a meaningful impact for the lowest students.

Results Regarding Student Perspectives

Students expressed a range of perspectives on the impact of the increased class size. Many students indicated that the level of support they felt decreased with the larger class. Surveys taken at the end of the year showed that students had a strong preference for the smaller class size with 65% of survey takers preferring to keep the class at its original size. Of the 35% who felt the larger class size was more beneficial, the emergent reasons were the additional friends that were accessible to them and the diversity of perspectives available. Among both groups a recurring pattern in the interviews was the change in student focus with the larger group

in addition to the shift in support available. One student reflected on the smaller class size, stating “it was easier to focus with no one off task distracting me. It was easier to learn.”

CHAPTER FIVE

CLAIM, EVIDENCE, REASONING

Claims from the Study

As evidenced by the stunningly similar performance metrics between the small group one teacher phase and the large group one teacher phase, class size alone had a minimal impact on overall student achievement. A study by Dee and West (see Chingos, 2011) referenced in a report published by the Brookings Institute echoed a similar finding for 8th grade students in their study: they found no overall impact on student test scores associated with class size. The researcher did not adjust the techniques of instruction. A 5E model with direct instruction components and guided inquiry was followed in all phases of the treatment.

However, the data did show a drop in performance across all quartiles in the time of transition to the larger class group- the large group with student teacher phase. In this phase, the mean score for the bottom quartile dropped to 69% and performance showed the greatest variability as indicated by the standard deviation of 0.22. Student interviews suggested that the students not only had to adjust to the larger class sizes, but also to the integration of new classmates, possible distractions, and shifting classroom routines. Many student interviews (55.6%) mentioned decreased focus as a concern impacting their achievement once the larger group was present. One student remarked, “You (the student) had to wait a bit longer when you had your hand raised. I got distracted when you (the teacher) were with another student and when I was waiting, I would talk to others and get sidetracked.” The academic drops seen in this

phase, but not in the settled larger group with one experienced teacher phase, reiterate that stable classroom routines and effective teacher leadership can have a profound impact on student achievement. Multiple studies corroborate this claim that the "most dominant factor affecting student academic gain is teacher effectiveness" (Colorado Department of Education, 2024, paragraph 5).

While class size may not have had as big of an impact as expected on achievement, this study did demonstrate the importance of the quality and consistency of our academic environments on achievement. Additional studies should investigate whether the grade decrease during the transition to larger classes could be better mitigated through targeted interventions like small group tutoring for the bottom 25% or direct teaching of routines for consistency.

Value of the Study and Consideration for Future Research

My study adds to a field of research that currently hosts many conflicting conclusions. It adds to the field creating additional corroboration for certain claims. The achievement data I've collected highlights the continued need for differentiation and targeted intervention for middle and bottom learners. As a teacher facing a possible transition to a larger group again this year due to short staffing, I know that I will want to do smaller breakout sessions, targeted frequent check-ins, and direct teaching of routines. Research consistently shows that such interventions are key for closing the achievement gap. John Hattie's work on visualizing learning through effect size identifies formative feedback and small-group interventions as having high effect sizes (small-group instruction at 0.47 effect size) and, therefore, as being well suited to help fill in the gaps for students who are less resilient in larger learning environments (Hattie, 2023). The

large class size did not dramatically alter class averages, but it was felt deeply by individuals on the margins of typical academic achievement, and I would aim to address that as a teacher. The rest of my research does demonstrate differences in work completion and significant changes in teacher burnout. Future research would be useful to find the balance between class size, maximum academic achievement, consistent student engagement, and stable teacher performance and health. It would also be useful to examine the possible compounding effect of large class sizes on teacher burnout over an extended time.

Impact on the Author

I am facing a similar situation again this school year, as we had a team member who found a job an hour closer to home. We were happy for her, but disheartened to hear we may need to increase our student load significantly again. The first time this happened, it was a whirlwind. I did not understand who would be most affected in the classroom, what should be a priority, or how it might impact me. I felt confused and overwhelmed.

Now, I have a significant knowledge base. I know that I should advocate against increasing the load on a fresh, first-year teacher that is now on our team as the research does not support that decision. I know that if the students are combined again, I should have interventions in place for the lowest students and should explicitly teach routines when the new students join. I would not worry about how this would impact academic achievement as much and instead would prioritize making students feel comfortable in the larger group and decreasing distractions. This process has given me a lot more power to advocate for and address the inequalities created by

larger class sizes. I am grateful to have chosen this topic and to have done the hard work to understand it better.

REFERENCES CITED

Abazaoglu, I., & Aztekin, S. (2016). The role of teacher morale and motivation on students' science and math achievement: Findings from Singapore, Japan, Finland, and Turkey. *Universal Journal of Educational Research, 4*(11), 2606-2617. https://eric.ed.gov/?id=EJ1118703

Blatchford, P., & Russell, A. (2020). Rethinking class size: The complex story of impact on teaching and learning. UCL Press. https://discovery.ucl.ac.uk/id/eprint/10112837

Brookhart, S. M. (2015). Graded Achievement, Tested Achievement, and Validity. *Educational Assessment*, 20(4), 268–296. https://doi.org/10.1080/10627197.2015.1093928

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology.* 3(2). 77-101. https://doi.org/10.1191/1478088706qp063oa

Chingos, M. M. (2011, May 11). *Class size: What research says and what it means for state policy.* Brookings Institution. https://www.brookings.edu/articles/class-size-what-research-says-and-what-it-means-for-state-policy/

Chingos, M.M. (2013), Class Size and Student Outcomes: Research and Policy Implications. *Journal of Policy Analysis and Management*, 32: 411-438. https://doi.org/10.1002/pam.21677

Colorado Department of Education. (n.d.). *6.2 The research: Effective teachers and the science of reading.* Retrieved November 23, 2024, from https://www.cde.state.co.us/cdesped/dyslexia-theresearch

Doan, S., Steiner, E., & Pandey, R. (2024, June 18). *Teacher well-being and intentions to leave in 2024: Findings from the 2024 State of the American Teacher Survey*. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1108-12.html

Filges, T., Sonne-Schmidt, C. S., & Viinholt Nielsen, B. C. (2018). *Small class sizes for improving student achievement in primary and secondary schools: A systematic review*. Campbell Collaboration.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's Class Size Study: Findings, Implications, Misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), 97–109. https://doi.org/10.2307/1164294

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "Why's" of Class Size: Student Behavior in Small Classes. *Review of Educational Research*, *73*(3), 321-368. https://doi.org/10.3102/00346543073003321

Gorard, S. and See, B.H. (2011), How can we enhance enjoyment of secondary school? The student view. *British Educational Research Journal*, 37: 671-690. https://doi.org/10.1080/01411926.2010.488718

Hattie, J. (2023). Hattie ranking: Influences and effect sizes related to student achievement. *Visible Learning*. https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/

Hanushek, E. A. (1999). *Evidence on class size*. Stanford University. Retrieved October 19, 2024, from https://hanushek.stanford.edu/sites/default/files/publications/Hanushek%201999%20EvidenceonCLassSize.pdf

Herman, K. C., Reinke, W. M., Dong, N., & Bradshaw, C. P. (2020, November 5). Can Effective Classroom Behavior Management Increase Student Achievement in Middle School? Findings From a Group Randomized Trial. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000641

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the Schools, 43*(6), 701–712. https://doi.org/10.1002/pits.20177

Levine, J. M., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 415–469). McGraw-Hill.

Lochmiller, C. R., Perrone, F., & Finley, C. (2024). Understanding school leadership's influence on teacher retention in high-poverty settings: An exploratory study in the U.S. *Education Sciences, 14*(5), 545. https://doi.org/10.3390/educsci14050545

Manlongat, M. D., Castor, A. D., De Chavez, R. B., Abila, R. D., Festijo, I. S., Fajilan, T. F., & Zuela, J. E. (2021). Impact of large class size on teachers' emotional and physical conditions. *International Multidisciplinary Research Journal, 3*(4), 11–21. https://doi.org/10.54476/iimrj230

Pennsylvania School Boards Association. (2024). *2024 State of Education Report* (p. 12). https://www.psba.org/wp-content/uploads/2024/01/2024-State-of-Education-report.pdf

Seidman, S. A., & Zager, J. (1986-1987). The Teacher Burnout Scale. *Educational Research Quarterly, 11*(1), 26–33.

Saloviita, T., & Pakarinen, E. (2021). Teacher burnout explained: Teacher-, student-, and organization-level variables. *Teaching and Teacher Education, 97*, 103221. https://doi.org/10.1016/j.tate.2020.103221

U.S. News & World Report. (n.d.). *Dover Area School District*. Retrieved January 19, 2025, from https://www.usnews.com/education/k12/pennsylvania/districts/dover-area-sd-102064

Wang, L., & Calvano, L. (2022). Class size, student behaviors, and educational outcomes. *Organization Management Journal, 19*(4), 126–142. https://doi.org/10.1108/OMJ-01-2021-1139

APPENDICES

APPENDIX A

IRB APPROVAL

IRB Training Complete:

Montana State University

ID 1096

IRB Social and Behavioral Research

Name	Stage	Completion Date	Expiration Date	Record ID
Rachael Gradwell	1 - Basic Course	15-Feb-2021	15-Feb-2026	40988263

From: NoReply@TOPAZTI.com <NoReply@TOPAZTI.com>
Sent: Friday, March 22, 2024 1:15 PM
To: Graves, John <graves@montana.edu>; Gradwell, Rachael <rachael.gradwell@student.montana.edu>; Szott, Brenna <brenna.szott@montana.edu>
Subject: IRB Protocol #2024-1424-EXEMPT APPROVED

External Sender

Hello Gradwell, Rachael,

Your protocol was reviewed by the IRB and has been approved.

PI: Gradwell, Rachael
Approval Date: 3/22/2024
Title: Impact of Classroom Size and Student Teacher Presence on Achievement, Work Completion, and Behavior in Middle School Students: A Comparative Study

Protocol #: 2024-1424-EXEMPT
Review Type: Exemption
Expiration Date: 3/22/2029

Work described under this protocol may now commence. The PI is responsible for ensuring that the protocol accurately describes research practices being conducted.

> Review Category designation determined by the IRB can be found in the final section of your protocol.
> Any changes must be submitted via Amendment prior to implementation.
> Per the Common Rule, research only requires Interim (annual) Review by the IRB if 1) it was reviewed via Full Committee or 2) is regulated by the FDA.
> All research is subject to post approval monitoring.
> All protocol types must be renewed 5 years after approval.
> Inform the IRB once your research is complete so that the protocol may be inactivated.

Please contact your IRB Program Manager with any questions or if you are in need of assistance. Thank you for your diligence in the care of human subjects research participants.

Institutional Review Board for the Protection of Human Subjects | Office of Research Compliance | Montana State University

Access your protocol anytime at https://montanaprod.topazti.net//Elements?emailLink=11%2c102%2c11437.

APPENDIX B

TEACHER BURNOUT SCALE

Teacher Burnout Scale

Seidman, Steven & Zager, Joanne. (1970). The Teacher Burnout Scale.. Educational Research Quarterly. 11. 26-33.

	Strongly Disagree	Disagree	Somewhat	Agree	Strongly Agree
I. Career Satisfaction					
1. I look forward to teaching in the future.					
2. I am glad I selected teaching as a career.					
3. Teaching is more fulfilling than I expected.					
4. If I had to do it all over again, I would not become a school teacher.					
5. I look forward to each teaching day.					
II. Perceived Administration Support					
6. I get adequate praise from my supervisors for a job well done.					
7. I feel my administrators are willing to help me with classroom problems, should they arise.					
8. I believe my efforts in the classroom are underappreciated by the administrators.					
9. My supervisors give me more criticism than praise.					
10. I feel administrators will not help me classroom difficulties.					
11. The administration blames me for					

classroom problems.					
III. Coping with Job-Related Stress					
12. I feel depressed because of my teaching experiences.					
13. The teaching day seems to drag on.					
14. My physical illness may be related to the stress of this job.					
15. I find it difficult to calm down after a day of teaching.					
16. I feel I could do a much better job at teaching if only the problems confronting me were not so great.					
17. The stresses of this job are more than I can bear.					
IV. Attitudes Towards Students					
18. The students act like a bunch of animals.					
19. Most of my students are decent people.					
20. Most students come to school ready to learn.					
21. Students come to school with bad attitudes.					

APPENDIX C

STUDENT INTERVIEW QUESTIONS

Student Interview Questions

1. What changes did you notice in classmates or the teacher support after the classes were combined?

    Probe: How did that change affect your learning or participation in class?

2. How did you feel about asking questions during class throughout the year? Did it change at all with the larger class sizes?

    Probe: Can you explain why you felt more/less comfortable asking questions in a larger class?

    Probe 2: Did the teacher response seem to change?

    Further Probe: Did you feel like that affected your understanding of class concepts?

3. If you could have kept our class size small like in the beginning of the year, would you? Why or why not?

    Probe: Do you think other students would agree with your opinion? Why/why not?