Gender Differences in Student Misbehaviors and Teacher Responses: Comparing Classrooms with Novice and Experienced Teachers

by

Eric A. Peist

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of Bachelor of Arts

With Honors in Psychology from the

University of Michigan

2013

Advisor:

Dr. Kai Cortina

Abstract

Gender bias in the classroom is a topic that has received significant attention, but empirical findings are inconclusive. This study is based on a previous thesis that explored gender differences in student misconduct for classrooms taught by expert (experienced) teachers. Through the previous project, video from 25 classrooms was utilized to create a coding system that categorizes misbehaviors and teacher responses (D'Lasnow, 2011). The current study validates the coding system for novice teachers and compares gender differences in misconduct and teacher responses depending on teaching expertise. One expert teacher and one novice teacher were paired with each classroom, so each pair of teachers provides two separate lessons for the same students. In total, 1379 incidents of student misbehavior were recorded for novice teachers, compared to 621 incidents for expert teachers. Findings for novices confirmed D'Lasnow's observations with experts: Talking was the most common form of misconduct and more than two thirds of misconducts received no response from the teachers. The gender ratio for total number of misbehaviors was similar for novice and expert teachers (approximately 55% of misbehaviors committed by male students). Hierarchical loglinear analysis of the combined expert/novice data set suggests that students are more off-task for novice teachers and that novice teachers are more likely to address the whole class in response to individual misbehaviors.

Keywords: gender bias, student misconduct, novice/expert teachers, video study

Gender Differences in Student Misbehavior and Teacher Responses: Comparing Classrooms with Novice and Experienced Teachers

The concept of teacher gender bias has been studied in psychology for over fifty years, yet there is no clear consensus on the extent of this gender bias. Several empirical studies were published, but one in particular garnered media attention: Sadker and Sadker (1986) reported findings of the study conducted between 1980 and 1984. According to these authors, there are major disadvantages for girls in the classroom, as boys receive more attention from teachers than girls, dominate class discussion, and are given more opportunities to succeed. Through the U.S. education system, the authors conclude that girls are encouraged to be quiet and submissive in the classroom, to refrain from taking math and science courses, and to value characteristics like organization and appearance over creativity and intellect. Often unintentionally, teachers allow these gender biases to persist (Sadker & Sadker, 1994).

Swinson and Harrop's (2009) research on classrooms in the United Kingdom supports the idea that girls are disadvantaged in school, as they found that boys were more likely than girls to receive both verbal approval and disapproval from teachers. Male students were given more instructing and redirecting than female students, likely influenced by the fact that boys were found to be off-task more frequently than girls.

A meta-analysis of 32 studies on gender differences in teacher-student interactions provides evidence for male-dominated student-teacher interactions, particularly, negative interactions. Across studies, teachers appear to initiate more negative interactions with boys, while positive reactions are observed equally often with boys and girls (Jones & Dindia, 2004). The gender gap in negative interactions could potentially be driven by gender differences in student behavior, as specific (mis-)behaviors may be more likely to demand a response from the teacher. The current study hypothesizes that these types of student behaviors vary between genders.

An observational study on teacher-student relationships in American elementary schools also investigated the role of gender in classrooms. Unlike Sadker and Sadker's research from approximately the same time, no teacher bias towards a particular gender was identified, as male and female students received similar amounts of positive interactions with teachers (e.g., compliments, praise, and teaching). Despite not reporting a gender bias, the results indicate that teachers are more likely to describe girls as obedient and boys as disruptive (Stake & Katz, 1982). Consistent with findings from Jones and Dindia (2004), both male and female teachers reprimanded boys more than girls (Stake & Katz, 1982).

Recent literature demonstrates that teacher gender has insignificant effects on the student gender bias. According to an observational study of three Canadian school districts by Duffy, Warren, and Walsh (2002), both male and female teachers generally tend to interact more with male students than female students. The study illustrated that this gender difference was not related to male students initiating conversations and interactions with teachers more often than female students. Another study from primary schools in the Netherlands observed that male and female teachers are both more likely to report relationships that involve conflict with male students (Split, Koomen, & Jak, 2012). These studies provide support for the claim that teacher gender does not have major effects on interactions with students, as male students receive more attention and appear more likely to have negative relationships with male and female teachers alike.

The Present Study

In the current study, teachers' experience status (novice/expert) will be explored to determine what role this variable has with respect to student gender interactions. Throughout the paper, "novice teachers" will refer to student teachers and "expert teachers" will refer to teachers who have their own classrooms and have been teaching for multiple years. For the novice-expert distinction, research shows experienced teachers have a particular advantage in classroom management. According to Ritter and Hancock (2004), teachers with less experience exercise more control over activities and student behaviors than teachers with many years of experience. The level of control exerted on a classroom is potentially related to the amount of interactions a teacher will initiate with students. While there has been significant research on both gender bias in the classroom and differences between novice and experienced teachers, the topics have not been studied simultaneously, and it is unknown, for example, whether novice teachers are more (or less) prone to gender bias. For this reason, the thesis will look at student misbehavior, teacher responses, student gender, and teacher experience. Student misbehaviors and teacher responses will be categorized according to a coding system developed for a previous project (D'Lasnow, 2011).

Because student teachers are newer to a classroom, students may be more likely to misbehave and "take advantage" of a new teacher. Research demonstrates that male students are more likely to be off-task and have greater levels of misconduct, and that teachers with less experience are prone to exerting more control over student behavior (Swinson & Harrop, 2009; Ritter & Hancock, 2004). This leads to the expectation that compared to experts, novices show a larger number of teacher to male student interactions. However, there is a possibility that student teachers demonstrate teaching styles that produce more gender equality, as they have more recently completed their training. Because of the attention that gender issues receive in higher education today, they may be better equipped to prevent gender bias in their interactions with students. This suggests that novice teachers are confronted with more misbehavior but tend to respond irrespective of the gender of the student.

While it is likely that novice teachers encounter student misbehavior more frequently than experts, there is no reason to believe that the distinction across the categories suggested by D'Lasnow (2011) differs depending on expertise. Overall, this study expects that boys commit more misbehaviors than girls, but that the distribution for each gender would not significantly alter between novices' and experts' classrooms.

The objectives/hypotheses of this study are:

1) To validate the coding scheme used by D'Lasnow (2011) in a sample of novice teachers. In his thesis, D'Lasnow only explored gender differences in student misbehaviors for expert teachers. It is predicted that the coding scheme is appropriate for novice teachers as well.

2) To see if patterns of student misbehavior are different for novice and expert teachers:

a) Do different kinds of misbehaviors occur at the same rate for novice and expert teachers?

b) Do expert and novice teachers differ in the way they respond to certain types of student misbehaviors?

c) Are gender differences more or less pronounced with novice teachers?

Methods

Sample

This thesis involves a secondary analysis of videos recorded in 2009-10 in 26 K-11 Southeast Michigan classrooms. These videos were originally collected for a study on attention, specifically using eye-tracking technology (Cortina, Miller, & Fang, in preparation). Because of the nature of the technology, each teacher was asked to wear glasses that were able to track eye movements, as well as a waist pack that contained a computer for data storage. The current study does not focus on the eye-tracking footage but video footage from stationary cameras placed around the classroom. The eye-tracking apparatus did not seem to affect teachers' lessons or interactions between teachers and students after the glasses were placed correctly on the teacher.

The videos are from schools located in three districts, covering a wide range of neighborhoods and socioeconomic statuses. In every classroom, two lessons were recorded on different days of the school year, one by a novice teacher and one by an expert teacher. Therefore, the students were essentially the same for each classroom's novice and expert lesson (aside from normal student absences). According to D'Lasnow (2011), the gender ratio for the coded expert lessons was approximately 1:1, as he reported 51% girls to 49% boys. Because the novice lessons were videotaped in the same classrooms, we can assume that the ratio will be the same for the novice lessons. In addition to the novice/expert distinction, teachers' lessons varied by school subject (e.g., math, science, English, and social studies) and grade level.

D'Lasnow (2011) coded 25 expert teacher videos for his thesis. The remaining videos were coded specifically for this study (26 novice teacher videos and 1 expert teacher video). One expert teacher video was coded for this study because it was not available when D'Lasnow analyzed the data.

The Videotaping Setup

All classrooms in the study had three stationary video cameras placed in different corners of the classroom, so every student was recorded by at least one of these cameras. The camera that provided views of the most students was utilized as the main camera for assessment of the classroom. If an incident was addressed by the teacher or heard on video but not visually captured on the main camera, footage from another camera that did cover the incident was assessed. Occasionally, video from other cameras was used to provide better angles of a situation. During some lessons, activities changed and students moved to different areas of the classroom, so in these instances, researchers switched the main camera.

Coding Scheme

This study employed D'Lasnow's (2011) coding system to determine if it applied to the videos of novice teachers as well as expert teachers. Coders recorded student misbehaviors, the gender of the student who committed the misbehavior, and the response from the teacher. The times of misbehaviors in the videos were also noted, so others could easily find all coded misbehaviors.

The coding system has eight categories for student misbehavior, which include disruptive hand raising (HR), calling out (CO), talking (T), gesturing (G), fidgeting (F), making noise (N), being off-task (OT), and other (O). These types of incidents were considered to be the most common forms of misbehavior in the classroom (D'Lasnow, 2011). Some of these categories represent normal student behavior, and would not be considered misconduct in every situation they occur. A misbehavior was only recorded when an action broke traditional United States classroom norms, affected a lesson, or interfered with learning. Disruptive hand raising (HR) involved a student raising his or her hand in a way that was distracting, usually with the intent to receive extra teacher attention. Calling out (CO) was coded when a student gave an answer or said something inappropriate out loud when he or she was not supposed to speak. When students were talking or laughing to each other in a way that was disruptive to the lesson or other students' learning, the incident was coded as talking (T). Gesturing (G) referred to disruptive movements meant to express meaning, while fidgeting (F) involved general disruptive movements. Making noise (N) was coded when students made distracting noises other than talking or laughing, and off-task (OT) was recorded when students were out of position in the classroom, doing something other than the assigned task, or not focusing on the lesson. Other (O) was coded for any other form of misbehavior not included in the previous seven categories, such as pushing or using physical force, copying another student's answers, or taking other students' learning materials (i.e. books, pencils, folders).

In some instances, a misbehavior could be included in more than one of the listed categories. When this situation arose, the most salient of the categories was coded. For example, if a student was banging his fists loudly onto his desk, but also was not completing the assigned work from the teacher, the coder would only code the misbehavior as "N" and not "OT" due to the clearly disruptive noise. There were other situations in which multiple students committed one misbehavior (e.g., three students were talking to each other). In this scenario, we recorded one misbehavior if the multiple students were the same gender, but two misbehaviors if both genders were involved. Because we are exploring gender differences, it was important that gender be taken into account for any specific misbehavior.

The coding system also includes six categories for teacher responses, which consist of reprimanding (R), saying a student's name (S), positively responding (P), producing a nonverbal reaction (NV), addressing the class (C), and no response (None) (see D'Lasnow, 2011). A teacher was recorded as reprimanding (R) when he or she verbally addressed a student's misbehavior. When the teacher only mentioned the student's name, the response was recorded as an "S." Positive responses (P) encouraged more misbehavior and were extremely rare for both novice and expert teachers. A nonverbal reaction (NV) occurred when a teacher addressed a misbehavior without using words, such as glaring at a student, and addressing the class (C) occurred when a teacher made an announcement to the whole classroom in response to a misbehavior. There were multiple incidents in which a teacher made one response for two misbehaviors occurring at approximately the same time. In this scenario, the same response was recorded for both misbehaviors.

The inter-rater reliability of the scheme was tested in D'Lasnow's thesis (2011), as two researchers coded a couple of the expert videos. A Cohen's Kappa of .76 was reported for student misbehaviors and a Cohen's Kappa of .82 was reported for teacher responses when both researchers coded the same incidents. Despite moderately high reliability on events that both coders recognized and similar numbers of misbehaviors and teacher responses, about 50% of the incidents each researcher coded were different from one another. These discrepancies in coding most likely came from moments in the videos when multiple misbehaviors were occurring at the same time (D'Lasnow, 2011).

Results

There were 2000 total misbehaviors for both novice and expert teachers. One thousand three hundred and seventy-nine misbehaviors were coded for the novice teachers' lessons (68.95%) and 621 misbehaviors were coded for the expert teachers' lessons (31.05%). Of the 2000 misbehaviors, 1463 received no response from the teacher (73.2%) and 537 received some form of teacher response (26.8%). One thousand one hundred and ten of the total number of misbehaviors were coded for male students (55.5%) and 890 misbehaviors were coded for female students (44.5%). Talking (42.8%) and off-task behaviors (20.0%) were the most common forms of misconduct.

Utilizing hierarchical loglinear analysis, a model was created to predict the frequency counts for a multidimensional contingency table with student gender, type of student misbehavior, teacher experience, and type of teacher response to misbehavior as predictor variables. Starting from the trivial complete model with all main effects and interactions with a $\chi^2(0) = 0$, hierarchical strategy was applied to find a parsimonious model. In a first step, the need to allow for a four-way interaction was ruled out; excluding this effect did not lead to a significant χ^2 test, indicating that the model implied expected frequency deviated from the observed counts only within the margins of error, χ^2 (6, N = 1975 incidents) = 4.505, $p = .609$ (See Table 1a).

We further reduced the complexity of the model by identifying a parsimonious generating model with insignificant χ^2. The final model contained one three-way interaction and all subordinated two-way interactions and main effects plus two separate interactions (student misbehavior x teacher response x teacher experience, gender x student misbehavior, & gender x teacher response). Due to the hierarchical nature of the analysis, all main effects were included in the model. This was expected since the marginal distributions were not expected to be equal for any of the independent variables. We eliminated the "other" category of student misbehavior because this category created a large amount of expected cells smaller than n = 5, which weakens the distributional properties of the test statistic. Because some of the categories for the teacher response variable also created expected cells smaller than n = 5, this variable was collapsed into two categories: "yes" for when the teacher made a response to an incident and "no" for when the teacher did not make a response to an incident (Tables 1a & 1b).

As an addition to the quality of the final model, we tested the distribution of the residuals for normal distribution. A Kolmogorov-Smirnov test was insignificant ($p = .989$), which means that there is no statistical reason to reject the assumption of normality given the effect of the final model (See Table 1c).

In Table 2, the interaction between student misbehavior and gender is explored utilizing the combined data set of expert and novice teacher lessons, χ^2 (6, N = 1975) = 35.043, $p < .001$ (See Table 2b). Results support the gender distribution of misbehaviors reported in D'Lasnow's (2011) research, as girls were more likely to perpetrate misconducts that fell into the hand raising and talking categories and boys were more apt to commit misbehaviors that fell into the off-task and gesture categories. The standard residuals (SR) for each cell indicate significant differences between boys and girls in these types of misbehaviors (SR > 2.0). Overall, more boys called out in class than girls, but the standard residuals were not as significant for calling out as they were for hand raising, talking, being off-task, and gesturing (See Table 2a).

Table 3 reveals the interaction between teacher response and gender, χ^2 (5, N = 2000) = 19.808, $p = .001$ (See Table 3b). It illustrates that girls are more likely than boys to receive no response from a teacher after a misbehavior (77.6% vs. 69.5%). Furthermore, boys receive more reprimands from teachers than girls (18.3% of boys' misbehaviors vs. 12.5% of girls' misbehaviors) (See Table 3a). These results are very much contingent on the type of misbehavior committed. Because talking and hand raising, categories that were rarely responded to by teachers, were girls' most common forms of misconduct, and off-task behavior, a category that was often responded to by teachers, was the boys' most common form of misconduct, the differences in non-responses from a teacher and reprimanding can be better explained. This interpretation is supported by the fact that the final model did not include the interaction type of student misbehavior x teacher response x student gender.

The interaction between student misbehavior, teacher response, and teacher experience was the only significant three-way interaction, χ^2 (5, N = 2000) = 275.531, $p < .001$ (See Table 4c). In order to understand the meaning of this effect, the two-way interactions of student response x teacher response were calculated separately for experts and novices and inspected for substantive differences (See Tables 4a & 4b). The three least common coded categories for misbehavior ("gesturing", "making noise", and "other") were combined so the expected counts of each cell were above 5 misbehaviors. The teacher response coding was also collapsed (response vs. no response) to avoid cell counts under n = 5.

The interaction between misbehavior and teacher response is represented through Table 5, χ^2 (7, N = 2000) = 298.416, $p < .001$ (See Table 5b). For this crosstabulation, the teacher response variable was collapsed into two categories again to avoid low expected cell counts. The misbehavior that most often received a reaction from the teacher was being off-task, as 56.2% of all misbehaviors coded as off-task were addressed by a teacher. Making noise and talking were the second and third most common misbehaviors to receive a teacher response (20.8% and 19.6% respectively). Hand raising, gesturing, and fidgeting were least likely to warrant responses (8.2%, 8.8%, & 9.6%) (See Table 5a).

Table 6 suggests that the distribution of responses towards misconduct for novice and expert teachers is different, χ^2 (5, N = 2000) = 34.759, $p < .001$ (See Table 6b). Novice teachers tend to address the class more than expert teachers, as 90.1% of all responses that addressed the class came from novices. Addressing the class accounted for 6.6% of responses for novice teachers and only 1.6% of responses for expert teachers. Novice teachers were also more likely than expert teachers to only say a student's name when reacting to a misbehavior, although this difference was not as pronounced as it was for addressing the class. Novice teachers did not respond to 71.4% of misconducts in the classroom while expert teachers did not respond to 77.1% of misconducts (See Table 6a).

The interaction between student misbehavior and teacher experience demonstrates differences in the types of misbehaviors committed in novice and expert classrooms, χ^2 (6, N = 1975) = 159.844, $p < .001$ (See Table 7b). Fidgeting, gesturing, and hand raising each accounted for a higher percentage of misbehaviors in expert classrooms (16.4%, 10.7%, & 11.2% respectively) than in novice classrooms (11.1%, 2.7%, & 3.1%), while being off-task accounted for a higher percentage of misbehaviors in novice classrooms (24.7% in novice classrooms vs. 10.2% in expert classrooms (See Table 7a).

The interaction between gender and teacher experience was not significant and therefore not an aspect of the loglinear model. Since there was no interaction between these two variables, it is inferred that the gender ratio of misbehaviors committed does not vary by novice or expert teachers. While boys are responsible for the majority of misconduct incidents in both types of teachers' classrooms, the percentage of male misbehaviors is approximately the same across teacher experience.

Discussion

Based on the recorded data, the distribution of male and female misbehaviors is similar regardless of teacher experience, supporting the hypothesis that the classroom gender bias for misconduct was close to the same for novice and expert teachers. Even though there was no clear gender discrepancy between classrooms with teachers of different experience levels, misbehaviors in novice and expert teachers' classrooms varied in other respects. Despite the virtually identical gender ratios for misbehaviors in both types of classrooms, in absolute terms, the novice teachers' classrooms demonstrated higher rates of misbehavior for boys and girls. An increased level of misconduct by students with novice teachers indicates that these teachers may exert less control over their students than expert teachers in general. This could represent a trend that may only be altered individually with years of real teaching practice, but this could also suggest a need for teacher preparation programs to better focus on methods of handling student misconduct.

The current analysis supports the claim that the types of misconduct occurring in the classrooms as well as some methods for responding to these behaviors are different for novice and expert teachers. For instance, off-task behavior that includes daydreaming, working on something other than assigned work, and not sitting at one's desk are more common forms of misbehavior with novice teachers than expert teachers. This type of misbehavior indicates disinterest in the lesson, as students are having trouble focusing on the assigned work. These results suggest that expert teachers are more effective at providing engaging lessons than student teachers. Whereas off-task behavior is more common in novices' classrooms, fidgeting and hand raising occur with higher likelihood in experts' classrooms than in inexperienced teachers' classrooms. Fidgeting and disruptive hand raising may illustrate enthusiasm about the content, as students are over-enthusiastic to answer questions or to see what will follow in the lesson. The largest inconsistency in teacher responses is seen in the number of times novices and experts addressed the class. Novice teachers addressed the class about nine times more than experts. Because of the larger number of misbehaviors in novice classrooms, reactions to misbehaviors that attend to the entire class may be less effective in eliminating unwanted behaviors on the part of the particular student that triggers the teacher response.

The percentage of misbehaviors not addressed by the teacher was fairly consistent across experience, as the majority of student misconduct incidents were not followed by responses. In many cases, not responding to a misbehavior can be an effective teaching strategy for limiting these actions in the future, as students do not receive extra attention for their disruptions. Other non-responses can be related to a teacher not being aware of its occurrence. Therefore, it is difficult to determine the true meaning of the reported non-responses by teachers. Future research could use the available eye-tracking data to examine whether a teacher is actually witnessing a misbehavior and ignoring it or missing the misconduct altogether. These two possible explanations for the large number of non-responses have different implications. If teachers are missing many incidents of misconduct, it could be important for teachers to receive more in-depth training on how to identify and handle misbehaviors, but if teachers are ignoring misconduct as an effective strategy to limit misbehaviors, it might be an indicator of expertise.

Evidence from this research also supports previously reported trends between gender and types of misbehavior committed in the classroom (D'Lasnow, 2011). Female students' misbehaviors included higher percentages of talking and hand raising, while male students' misbehaviors consisted of higher percentages of off-task incidents. Because teachers were prone to responding to off-task behavior in the classroom more than any other type of misconduct, this is consistent with the results that show boys received more reprimands and girls received more "non-responses" from the teachers. Off-task behaviors can often be more disruptive to a lesson than talking or hand raising, and the gender bias in terms of how teachers respond may be explained through this difference.

Limitations

One limitation of this study is a potential coder effect that is confounded with the effect of expertise. There were only two coders for the combined data set (one for expert teachers and one for novice teachers). The second coder learned the system by first coding videos that

D'Lasnow (2011) already completed. Through the whole coding process, it was revealed that the coder of the novice videos had a lower threshold to code misbehaviors than D'Lasnow, resulting in a higher count of incidents for the novice teachers' classrooms. Future studies should utilize more trained coders who work through sample videos together before starting their own separate coding. With more coders and a thorough training process, the current system would be more effective, as issues such as coder sensitivity, when to break up incidents into multiple misconducts, and how to code multiple misbehaviors at the same time could be standardized.

While the coder discrepancy between the novice and expert videos exists and may exaggerate the number of misbehaviors occurring in novice classrooms, a significant difference in classroom conduct remains. To determine the extent of the coder effect, the novice video coder, examined three expert videos completed by D'Lasnow (2011) at random. In these videos, the novice coder reported 1.6 misbehaviors for every misbehavior reported by the expert coder. Utilizing this ratio, if there was no effect between expert and novice teachers, it would be expected that the coder discrepancy leads to approximately 994 misbehaviors in the novice videos (621 coded misbehaviors in expert classrooms x 1.6 = 993.6 expected misbehaviors in novice teacher classrooms). Because 1379 misbehaviors were coded in novice classrooms, a substantial difference in number of misconducts between novice and expert teacher classrooms remains even after accounting for the coder effect. Although no statistical test could be performed, we conclude that there is, in fact, a higher incidence rate with novice teachers.

Another limitation of the empirical paradigm used in both studies is the inability to determine from the code which individual committed each misbehavior. Some classrooms were well behaved except for one or two students, and these students accounted for a substantial share of misbehaviors observed. Future research could consider coding for specific students, as some outliers may contribute to the overall results.

An additional limitation of this study is that the age range and grade level of each classroom is not factored into the analysis. For example, fidgeting often appeared to be more common with younger children. However, including more independent variables would make the test statistics of the loglinear model in the current study questionable due to the increased number of cells with an expected cell count of less than $n = 5$. Including more variables in the model would require a larger sample size. Another limitation of the study is the fact that observations of student misbehaviors are not fully independent, as students are clustered within classrooms. One misbehavior may trigger another, or one specific teacher may be more prone to allowing misbehaviors in his or her classroom. This was ignored in the statistical analysis and might have inflated the degrees of freedom.

Finally, the videos utilized in the study are all from one area of Michigan. While the data comes from a variety of school districts with a large range of socioeconomic statuses, research in other areas of the U.S. should be explored to determine if the same classroom trends exist. Furthermore, the study used a convenience sample in which teachers agreed to participate in an extensive classroom study. Because of the nature of the research, it is possible that teachers with difficult classrooms refused to participate, leaving out the classrooms with the most disruptive students.

Conclusion

This study provides evidence for gender differences in types of classroom misconduct, differences in the types of misbehaviors committed in classrooms under expert and novice teacher supervision, and differences in teacher responses to misconduct based on teacher experience. Future research should attempt to gain a deeper understanding of gender misbehavior patterns in the classroom. For example, why are girls more likely to talk than be off-task? Disparity between novice and expert teachers' lessons and teaching strategies should also be explored in more aspects than classroom misbehavior. Variation in student misbehavior and subsequent responses for novices and experts likely means that experience level affects the classroom in multiple ways. How do students' achievement, interest in the curriculum, and classroom participation vary by teacher experience? While this thesis contributes to research on gender bias in the classroom, effects of teacher experience on students, and classroom misconduct, potential studies should further examine these aspects of education and how they relate to one another.

References

Cortina, K., Miller, K. F., & Fang, G. (In preparation). *Measuring teacher attention using mobile eye tracking technique*.

D'Lasnow, R. D. (2011). *Do boys really get away with "calling out?" Testing the gender bias hypothesis for disruptive behavior in the classroom* (Unpublished honors thesis). University of Michigan, Ann Arbor, MI.

Duffy, J., Warren, K., & Walsh, M. (2002). Classroom interactions: Gender of teacher, gender of student, and classroom subject. *Sex Roles, 45*(9/10), 579-593.

Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research, 74*(4)*,* 443-471.

Ritter, J. T., & Hancock, D. R. (2004). Exploring the relationship between certification sources, experience levels, and classroom management orientations of classroom teachers. *Teaching and Teacher Education, 23,* 1206-1216.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How america's schools cheat girls.* New York: Macmillan Publishing Company.

Sadker, M., & Sadker, D. (1986). Sexism in the classroom: From grade school to graduate school. *The Phi Delta Kappan, 67*(7), 512-515.

Split, J. L., Koomen, H. M. Y., & Jak, S. (2012). Are boys better off with male and girls with female teachers? A multilevel investigation of measurement invariance and gender match in teacher-student relationship quality. *Journal of School Psychology, 50,* 363-378.

Stake, J. E., & Katz, J. F. (1982). Teacher-pupil relationships in the elementary school classroom: Teacher-gender and pupil-gender differences. *American Educational Research Journal, 19*(3)*,* 465-471.

Swinson, J., & Harrop, A. (2009). Teacher talk directed to boys and girls and its relationship to their behaviour. *Educational Studies*, *35*(5), 515-524.

Author Note

Eric A. Peist, Department of Psychology, University of Michigan, Ann Arbor

Thank you to Kai Cortina for collaborating with me on the project, and providing me with all of the guidance, support, and resources I needed to ensure the study was successful. It was a privilege to work with Kai and I appreciate all of the hard work he put into the project. I would also like to thank Kevin Miller and the Vizlab for their assistance throughout the process.

Table 1a

K-Way and Higher-Order Effects

	K	df	Likelihood Ratio		Pearson		Number of Iterations
			Chi-Square	*p*	Chi-Square	*p*	
K-way and Higher Order Effects[a]	1	55	2713.223	.000	3898.875	< .001	0
	2	46	539.978	.000	577.280	< .001	2
	3	25	57.406	.000	52.279	.001	5
	4	6	5.364	.498	4.505	.609	5
K-way Effects[b]	1	9	2173.245	.000	3321.595	< .001	0
	2	21	482.572	.000	525.001	< .001	0
	3	19	52.042	.000	47.775	< .001	0
	4	6	5.364	.498	4.505	.609	0

a. Tests that k-way and higher order effects are zero.
b. Tests that k-way effects are zero.

Table 1b

Goodness-of-Fit Tests

	Chi-Square	df	*p*
Likelihood Ratio	27.015	20	.135
Pearson	26.051	20	.164

Table 1c

One-Sample Kolmogorov-Smirnov Test

Null Hypothesis	*p*	Decision
The distribution of the standard residuals for the created loglinear model is normal with a mean of 0.01 and standard deviation 0.69	.989	Retain the null hypothesis.

Table 2a

*Student Misbehavior * Gender Crosstabulation*

			Gender		
			Female	Male	Total
Misbehavior	CO	Count	79	123	202
		% within Misbehavior	39.1%	60.9%	100%
		% within Gender	9.0%	11.2%	10.2%
		Std. Residual	-1.2	1.1	
	F	Count	111	140	251
		% within Misbehavior	44.2%	55.8%	100%
		% within Gender	12.6%	12.8%	12.7%
		Std. Residual	-.1	.1	
	G	Count	33	69	102
		% within Misbehavior	32.4%	67.6%	100%
		% within Gender	3.7%	6.3%	5.2%
		Std. Residual	-1.9	1.7	
	HR	Count	65	45	110
		% within Misbehavior	59.1%	40.9%	100%
		% within Gender	7.4%	4.1%	5.6%
		Std. Residual	2.3	-2.0	
	N	Count	20	33	53
		% within Misbehavior	37.7%	62.3%	100%
		% within Gender	2.3%	3.0%	2.7%
		Std. Residual	-.7	.7	
	OT	Count	150	250	400
		% within Misbehavior	37.5%	62.5%	100%
		% within Gender	17.0%	22.9%	20.3%
		Std. Residual	-2.1	1.9	
	T	Count	423	434	857
		% within Misbehavior	49.4%	50.6%	100%
		% within Gender	48.0%	39.7%	43.4%
		Std. Residual	2.1	-1.9	
Total		Count	881	1094	1975
		% within Misbehavior	44.6%	55.4%	100%
		% within Gender	100%	100%	100%

Table 2b

Chi-Square Tests

	Value	df	*p* (2-sided)
Pearson Chi-Square	35.043[a]	6	< .001
Likelihood Ratio	35.273	6	< .001
Linear-by-Linear Association	4.935	1	.026
No of Valid Cases	1975		

Table 3a

*Teacher Response * Gender Crosstabulation*

			Gender		
			Female	Male	Total
Teacher Response	None	Count	691	772	1463
		% within Teacher Response	47.2%	52.8%	100%
		% within Gender	77.6%	69.5%	73.2%
		Std. Residual	1.6	-1.4	
	C	Count	45	56	101
		% within Teacher Response	44.6%	55.4%	100%
		% within Gender	5.1%	5.0%	5.1%
		Std. Residual	.0	.0	
	NV	Count	8	15	23
		% within Teacher Response	34.8%	65.2%	100%
		% within Gender	0.9%	1.4%	1.2%
		Std. Residual	-.7	.6	
	P	Count	10	14	24
		% within Teacher Response	41.7%	58.3%	100%
		% within Gender	1.1%	1.3%	1.2%
		Std. Residual	-.2	.2	
	R	Count	111	203	314
		% within Teacher Response	35.4%	64.6%	100%
		% within Gender	12.5%	18.3%	15.7%
		Std. Residual	-2.4	2.2	
	S	Count	25	50	75
		% within Teacher Response	33.3%	66.7%	100%
		% within Gender	2.8%	4.5%	3.8%
		Std. Residual	-1.4	1.3	
Total		Count	890	1110	2000
		% within Teacher Response	44.5%	55.5%	100%
		% within Gender	100%	100%	100%

Table 3b

Chi-Square Tests

	Value	df	*p* (2-sided)
Pearson Chi-Square	19.808[a]	5	.001
Likelihood Ratio	20.123	5	.001
Linear-by-Linear Association	19.271	1	< .001
N of Valid Cases	2000		

a. 0 cells have expected count less than 5. The minimum expected count is 10.24.

Table 4a

*Student Misbehavior * Teacher Response Crosstabulation for Novice Teachers*

			Teacher Response		
			No	Yes	Total
Misbehavior	CO	Count	96	53	149
		% within Misbehavior	64.4%	35.6%	100%
		% within Teacher Response	9.8%	13.4%	10.8%
		Std. Residual	-1.0	1.6	
	OT	Count	163	175	338
		% within Misbehavior	48.2%	51.8%	100%
		% within Teacher Response	16.6%	44.3%	24.5%
		Std. Residual	-5.0	7.9	
	T	Count	484	124	608
		% within Misbehavior	79.6%	20.4%	100%
		% within Teacher Response	49.2%	31.4%	44.1%
		Std. Residual	2.4	-3.8	
	HR	Count	35	7	42
		% within Misbehavior	83.3%	16.7%	100%
		% within Teacher Response	3.6%	1.8%	3.0%
		Std. Residual	.9	-1.5	
	F	Count	136	15	151
		% within Misbehavior	90.1%	9.9%	100%
		% within Teacher Response	13.8%	3.8%	10.9%
		Std. Residual	2.7	-4.3	
	G+N+O*	Count	70	21	91
		% within Misbehavior	76.9%	23.1%	100%
		% within Teacher Response	7.1%	5.3%	6.6%
		Std. Residual	.6	-1.0	
Total		Count	984	395	1379
		% within Misbehavior	71.4%	28.6%	100%
		% within Teacher Response	100%	100%	100%

Note. *For the purpose of ensuring the expected value of every cell in this analysis was over 5, the three least common student misbehavior categories were combined to form a larger "other" category.

Table 4b

*Student Misbehavior * Teacher Response Crosstabulation for Expert Teachers*

			Teacher Response		
			No	Yes	Total
Misbehavior	CO	Count	29	24	53
		% within Misbehavior	54.7%	45.3%	100%
		% within Teacher Response	6.1%	16.9%	8.5%
		Std. Residual	-1.9	3.4	
	OT	Count	12	50	62
		% within Misbehavior	19.4%	80.6%	100%
		% within Teacher Response	2.5%	35.2%	10.0%
		Std. Residual	-5.2	9.5	
	T	Count	205	44	249
		% within Misbehavior	82.3%	17.7%	100%
		% within Teacher Response	42.8%	31.0%	40.1%
		Std. Residual	.9	-1.7	
	HR	Count	66	2	68
		% within Misbehavior	97.1%	2.9%	100%
		% within Teacher Response	13.8%	1.4%	11.0%
		Std. Residual	1.9	-3.4	
	F	Count	91	9	100
		% within Misbehavior	91.0%	9.0%	100%
		% within Teacher Response	19.0%	6.3%	16.1%
		Std. Residual	1.6	-2.9	
	G+N+O*	Count	76	13	89
		% within Misbehavior	85.4%	14.6%	100%
		% within Teacher Response	15.9%	9.2%	14.3%
		Std. Residual	.9	-1.6	
Total		Count	479	142	621
		% within Misbehavior	77.1%	22.9%	100%
		% within Teacher Response	100%	100%	100%

Note. *For the purpose of ensuring the expected value of every cell in this analysis was over 5, the three least common student misbehavior categories were combined to form a larger “other” category.

Table 4c

Chi-Square Tests

Teacher Experience		Value	df	*p* (2-sided)
Novice	Pearson Chi-Square	142.411[b]	5	< .001
	Likelihood Ratio	140.761	5	< .001
	Linear-by-Linear Association	62.817	1	< .001
	N of Valid Cases	1379		
Expert	Pearson Chi-Square	165.913[c]	5	< .001
	Likelihood Ratio	149.024	5	< .001
	Linear-by-Linear Association	64.494	1	< .001
	N of Valid Cases	621		
Total	Pearson Chi-Square	275.531[a]	5	< .001
	Likelihood Ratio	267.036	5	< .001
	Linear-by-Linear Association	128.805	1	< .001
	N of Valid Cases	2000		

a. 0 cells have expected count less than 5. The minimum expected count is 29.54.
b. 0 cells have expected count less than 5. The minimum expected count is 12.03.
c. 0 cells have expected count less than 5. The minimum expected count is 12.12.

Table 5a

*Student Misbehavior * Teacher Response Crosstabulation*

			Teacher Response		
			No	Yes	Total
Misbehavior	CO	Count	125	77	202
		% within Misbehavior	61.9%	38.1%	100%
		% within Teacher Response	8.5%	14.3%	10.1%
		Std. Residual	-1.9	3.1	
	F	Count	227	24	251
		% within Misbehavior	90.4%	9.6%	100%
		% within Teacher Response	15.5%	4.5%	12.6%
		Std. Residual	3.2	-5.3	
	G	Count	93	9	102
		% within Misbehavior	91.2%	8.8%	100%
		% within Teacher Response	6.4%	1.7%	5.1%
		Std. Residual	2.1	-3.5	
	HR	Count	101	9	110
		% within Misbehavior	91.8%	8.2%	100%
		% within Teacher Response	6.9%	1.7%	5.5%
		Std. Residual	2.3	-3.8	
	N	Count	42	11	53
		% within Misbehavior	79.2%	20.8%	100%
		% within Teacher Response	2.9%	2.0%	2.6%
		Std. Residual	.5	-.9	
	OT	Count	175	225	400
		% within Misbehavior	43.8%	56.2%	100%
		% within Teacher Response	12.0%	41.9%	20.0%
		Std. Residual	-6.9	11.3	
	T	Count	689	168	857
		% within Misbehavior	80.4%	19.6%	100%
		% within Teacher Response	47.1%	31.3%	42.8%
		Std. Residual	2.5	-4.1	
	O	Count	11	14	25
		% within Misbehavior	44.0%	56.0%	100%
		% within Teacher Response	0.8%	2.6%	1.2%
		Std. Residual	-1.7	2.8	
Total		Count	1463	537	2000
		% within Misbehavior	73.2%	26.8%	100%
		% within Teacher Response	100%	100%	100%

Table 5b

Chi-Square Tests

	Value	df	*p* (2-sided)
Pearson Chi-Square	298.416[a]	7	< .001
Likelihood Ratio	292.184	7	< .001
Linear-by-Linear Association	6.340	1	.012
N of Valid Cases	2000		

a. 0 cells have expected count less than 5. The minimum expected count 6.71.

Table 6a

*Teacher Response * Teacher Experience Crosstabulation*

			Teacher Experience		
			Novice	Expert	Total
Teacher Response	None	Count	984	479	1463
		% within Teacher Response	67.3%	32.7%	100%
		% within Teacher Experience	71.4%	77.1%	73.2%
		Std. Residual	-.8	1.2	
	C	Count	91	10	101
		% within Teacher Response	90.1%	9.9%	100%
		% within Teacher Experience	6.6%	1.6%	5.1%
		Std. Residual	2.6	-3.8	
	NV	Count	19	4	23
		% within Teacher Response	82.6%	17.4%	100%
		% within Teacher Experience	1.4%	0.6%	1.2%
		Std. Residual	.8	-1.2	
	P	Count	12	12	24
		% within Teacher Response	50.0%	50.0%	100%
		% within Teacher Experience	0.9%	1.9%	1.2%
		Std. Residual	-1.1	1.7	
	R	Count	212	102	314
		% within Teacher Response	67.5%	32.5%	100%
		% within Teacher Experience	15.4%	16.4%	15.7%
		Std. Residual	-.3	.5	
	S	Count	61	14	75
		% within Teacher Response	81.3%	18.7%	100%
		% within Teacher Experience	4.4%	2.3%	3.8%
		Std. Residual	1.3	-1.9	
Total		Count	1379	621	2000
		% within Teacher Response	69.0%	31.0%	100%
		% within Teacher Experience	100%	100%	100%

Table 6b

Chi-Square Tests

	Value	df	*p* (2-sided)
Pearson Chi-Square	34.759[a]	5	< .001
Likelihood Ratio	39.921	5	< .001
Linear-by-Linear Association	1.501	1	.220
N of Valid Cases	2000		

a. 0 cells have expected count less than 5. The minimum expected count is 7.14.

Table 7a

*Student Misbehavior * Teacher Experience Crosstabulation*

			Teacher Experience		
			Novice	Expert	Total
Misbehavior	CO	Count	149	53	202
		% within Misbehavior	73.8%	26.2%	100%
		% within Teacher Experience	10.9%	8.7%	10.2%
		Std. Residual	.8	-1.2	
	F	Count	151	100	251
		% within Misbehavior	60.2%	39.8%	100%
		% within Teacher Experience	11.1%	16.4%	12.7%
		Std. Residual	-1.7	2.6	
	G	Count	37	65	102
		% within Misbehavior	36.3%	63.7%	100%
		% within Teacher Experience	2.7%	10.7%	5.2%
		Std. Residual	-4.0	6.0	
	HR	Count	42	68	110
		% within Misbehavior	38.2%	61.8%	100%
		% within Teacher Experience	3.1%	11.2%	5.6%
		Std. Residual	-3.9	5.9	
	N	Count	41	12	53
		% within Misbehavior	77.4%	22.6%	100%
		% within Teacher Experience	3.0%	2.0%	2.7%
		Std. Residual	.7	-1.1	
	OT	Count	338	62	400
		% within Misbehavior	84.5%	15.5%	100%
		% within Teacher Experience	24.7%	10.2%	20.3%
		Std. Residual	3.7	-5.5	
	T	Count	608	249	857
		% within Misbehavior	70.9%	29.1%	100%
		% within Teacher Experience	44.5%	40.9%	43.4%
		Std. Residual	.6	-.9	
Total		Count	1366	609	1975
		% within Misbehavior	69.2%	30.8%	100%
		% within Teacher Experience	100%	100%	100%

Table 7b

Chi-Square Tests

	Value	df	*p* (2-sided)
Pearson Chi-Square	159.844[a]	6	< .001
Likelihood Ratio	155.623	6	< .001
Linear-by-Linear Association	23.910	1	< .001
N of Valid Cases	1975		

a. 0 Cells have expected count less than 5. The minimum expected count is 16.34.