Are Coeducational Classes Truly Coeducational?

Carol Lynn Martin, Sonya Xinyue Xiao, Richard A. Fabes, Laura D. Hanish, Dawn DeLay, &

Krista Oswalt

Arizona State University

Manuscript Accepted to *Elementary School Journal*

Date: August 11, 2021

Submitted ERIC Dec 22, 2022

Revision Submitted: July 28, 2022

Acknowledgements

We thank the superintendents, teachers, parents, and students in the Mesa and Kyrene Districts for their interest and cooperation with the research project. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180028 to C. Martin (PI) at Arizona State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

Boys and girls sit together in most classrooms, but do they interact? Based on forty-year-old evidence, coeducational classes may not be coeducational but instead be segregated by gender, which may undermine student success. Our goal is to answer this question in today's classrooms. We used longitudinal data to assess gender segregation in 26 classes in three U.S. coeducational elementary schools over an academic year. Third- to fifth-grade students (n = 515) from diverse backgrounds were asked how often they work with (frequency) and how well they work together (quality) with each classmate. Analyses illustrated a strikingly consistent pattern: For every grade, gender, and classroom, and across both fall and spring, students reported that they interacted more frequently with and had higher quality interactions with same- than with other-gender classmates.

Implications: Given the findings, teachers should encourage mixed-gender interactions; students likely will benefit socially and academically from these efforts.

Keywords Gender segregation, classroom community, gender, peer interactions, learning environment

Word count = 148

Are Coeducational Classes Truly Coeducational?

Most K-12 schools in the U.S. are coeducational with boys and girls sitting together in classrooms, but do they engage each other within their learning environments? Based on the available evidence, there is reason to question whether coeducational classes are actually coeducational. Forty years ago, researchers demonstrated that "sex segregation"—in which students spend more time interacting with same-gender peers and less with other-gender peers--was common in elementary school classrooms, and this kind of segregation resulted from and contributed to the formation of gender stereotypes (Lockheed, 1986; Lockheed & Harris, 1982, 1984; Tuma & Hallinan, 1979). Early studies of segregation in classrooms focused on the conditions in which small group instruction was used and then examined the nature of those groupings (Lockheed & Harris, 1982, 1984). In these studies, mixed-gender groupings (i.e., working with both girls and boys) were rare, accounting for only about 8% of the groupings. However, no contemporary research has assessed whether U.S. public school coeducational classrooms are segregated vs coeducational.

Gender segregation relates to a variety of social and academic outcomes as demonstrated in young children and in elementary-age children (Fabes et al., 2018; Maccoby, 1998). Gender-segregated peer interactions are related to increased stereotyping (Lockheed & Harris, 1984), gender-typical behavior (Fabes et al., 1997), stress when interacting with cross-gender peers, and behavioral issues (for boys) (Fabes et al., 1997). Further, the more children segregate, the stronger the consequences (Martin & Fabes, 2001). In contrast, gender-integrated classroom interactions can be beneficial for students to gain opportunities for peer learning and to form positive academically-related beliefs (Lockheed & Harris, 1982, 1984).

Because gender-segregation in classes can be problematic and gender-integration can be beneficial (Fabes et al., 2018; Lockheed, 1986; Lockheed & Harris, 1982; Martin et al., 2014), it is crucial to understand the degree to which today's elementary classrooms are gender segregated. To accomplish this, we asked students about their experiences working with each student in their class and used their responses to calculate and compare the frequency and quality of their interactions with same- and other-gender peers. Gender segregation (GS) would be considered to occur if the frequency and quality of same-gender interactions are significantly greater than other-gender interactions. Maximal gender integration (GI) would be considered to occur if the frequency and quality of same- and other-gender interactions are similar to one another, meaning students have a wide range of interactional opportunities, and that these opportunities encompass many students regardless of gender. Lower levels of GI might be seen if a student worked mainly with same-gender peers (frequency) and worked well (quality) primarily with same-gender students. If many students work mainly with same-gender students, this would necessarily constrain options for interactions with other-gender students for other classmates who may desire more integration although quality of interaction could still be a way to feel more integrated within the class. In the case of low levels of GI, the student's opportunities are limited compared to a student who works well with many same- and other-gender students. Thus, the more students gain opportunities with a wider range of students, the more exposure to diversity and to a wider range of experiences. In this way, we propose that increasing levels of GI interactions serve to broaden student participation and support, thereby promoting school success.

Gendered Relations in the Classroom

As noted, attendance in coeducational classes provides no assurance that boys and girls

work effectively together to learn, solve academic problems, and support one another in their academic efforts (Fabes et al., 2018). The same classes that appear on the surface to be integrated by gender may be quite segregated, meaning that most students interact and work with same- rather than with other-gender classmates. Because most adults assume that having boys and girls together in a classroom ensures that they work together, attention to GS in the classroom has gone relatively unstudied and is assumed to be unimportant. However, focus-group research hints that GS may still persist and needs to be studied in more depth. For example, in these focus groups, it was found that there were some young children in coeducational classrooms in which boys did not know the names of any of the girls in their class and in which students felt uncomfortable working with other-gender classmates. Additionally, many teachers reported that they were unaware of these facts (Miller & Updegraff, 2014)..

Relative to research on racial or ethnic segregation/integration and educational success, little attention has been paid to GS or how GI may improve academic outcomes for students. This is true despite the fact that Lockheed and colleagues identified this concern over 30 years ago. Many of the points Lockheed and Harris (1984, p. 276) noted likely still apply today. They argued that, unlike racial segregation, gender segregation is not seen as a problem. Specifically, they noted that "*concerns regarding the possible negative consequences of this kind of managerial strategy, or of sex segregation in general, typically are discounted by the belief that the separatist behaviors and attitudes of boys and girls represent a brief natural stage in their development rather than a lasting pattern of interaction and attitudes. Unfortunately, this line of reasoning masks a serious underlying problem. It appears that sex segregation may have social consequences that persist into adulthood."*

If gender is seen as a problem in classrooms, the issue tends to center on questions about

separating girls and boys into segregated classrooms/schools (Brown, 2013; Halpern et al., 2011). Furthermore, teachers may even use mixing of girls and boys as a punishment rather than to enhance learning (Lockheed & Harris, 1984). We contend that a more sensible and scalable model is to promote positive working relationships between male and female students in coeducational classrooms. But research is lacking on the degree to which male and female students actually work together within coeducational contexts; a gap that the present research is designed to address.

Factors Potentially Influencing Gender Segregation in Today's Classroom

Since the publication of the early studies on gender segregated classrooms, there have been substantial changes in gender relations in the U.S. For example, with increased numbers of women in the labor force and with advanced degrees, there have been documented changes in gender stereotypes, and changes in the perception of the competence of women being similar to that of men (Eagly et al., 2020). Given such societal shifts, it is reasonable to expect that the forces that contributed to gender segregation in classrooms are no longer evident in U.S. schools. However, there is also evidence from other settings that suggests that gender segregation is a strong preference (Mehta & Strough, 2009); thus, it may be continuing within schools.

Whereas there are some developmental differences to patterns of GS and GI, these are minimal compared to the tendencies seen across childhood and adolescence to segregate by gender when given the opportunity. Individuals tend to seek out others "like" themselves on a variety of dimensions (language, age, race/ethnicity) (McPherson et al., 2001). The tendency to segregate by gender is a very common pattern and has been claimed to be one of the most pervasive developmental phenomenon seen (Maccoby, 1998). Outside of school settings, a number of studies in the past 20 years have demonstrated the powerful tendency of children to

assort themselves by gender when given the opportunity (e.g., in sports, in social activities; for review see Mehta & Strough, 2009), and earlier studies suggested that these patterns are evident at a young age (La Freniere et al., 1984; Martin & Fabes, 2001; Martin et al., 1999; Serbin et al., 1994). This tendency strengthens considerably over early to middle childhood (Maccoby, 1998; Maccoby & Jacklin, 1987) and these same-gender preferences were evident in elementary classrooms in the past (Tuma & Hallinan, 1979; Webb, Baxter, Thompson, 1997). Empirical evidence of children's preferences for both familiar and unfamiliar same-gender interaction partners is strong, likely due to expecting and experiencing shared similarities with same gender peers (e.g., Martin, 1989; Martin et al., 2011; Zosuls et al., 2011). Even in adolescents, when heterosexual students show interest in other-gender peers, there continues to be strong segregation by gender, although more gender mixing in groups is apparent (Mehta & Strough, 2009; Strough, Swenson, & Cheng (2001).

Contexts also matter to the degree of GS observed: the more children have choices in their interaction partners, the more gender segregation is observed (Fabes et al., 2018). However, no recent research has gathered information about GS or GI in the classroom, so this question remains unanswered. The concern is that these early patterns of GS may have both short and long-term consequences, influencing how students learn and then how they relate to one another later in life in their personal relationships and in the workplace (Maccoby, 1998). These consequences increase the need for educators to attend to issues of GS and GI in their classrooms (Fabes et al., 2018).

Measuring Gender Segregation/Integration

There are a number of ways to assess GI versus GS in today's schools (Fabes et al., 2018), and each of these methods is informative. Measuring the frequency of interaction and

quality of interactions are two methods; each provides different insights into why segregation might be happening and how integration might be promoted. Frequency of interaction provides insights into the number of opportunities a student might have to interact with same- and other-gender classmates. Interactional quality concerns whether interactions with same- and other-gender classmates are effective and positive. Furthermore, comparing same- and other-gender classmate frequency and quality of interaction provides indications whether classroom interactions are more gender segregated (more frequent and higher quality of interactions in same- versus other-gender interactions) or are more gender integrated (similar frequency and quality of interactions across both types), as might be assumed given the coeducational setting.

Although we acknowledge that gender is better conceptualized as a spectrum than binary (Hyde et al., 2018), because most young students view themselves and others as falling into binary categories ("I'm a boy") (Glazier et al., 2020),which we focused our research on these categories. We consider same-gender peers to be those peers in the same gender group as the target student and other-gender peers to be those peers in the other gender group as the target student.

The Current Study

The primary goal of the present study is to assess the degree to which coeducational classes are actually coeducational. To achieve this goal, we examined 3rd to 5th grade students' perceptions of the amount of time spent working with their same- and other-gender classmates (i.e., interaction frequency), and their assessments of how well they work on class assignments with same- and other-gender classmates (i.e., interaction quality). Students in middle elementary school spend most of their day with the same classmates so this is an ideal age group to test for gender segregation/integration. Because data were collected at two times over the course of an

academic year (fall and spring), short-term changes over the academic year were also examined. Although we explored gender differences in students' perceptions of interaction quality and frequency with their classmates, we made no predictions about gender-related differences in patterns because both girls and boys tend to show strong and consistent levels of gender segregation (see Mehta & Strough, 2009). To provide additional information about classroom variation, we examined the descriptive information concerning patterns of gender segregation as they varied across classrooms, using class-averaged measures of frequency and quality of interaction with same- and other-gender students.

The first research question we addressed concerned interaction frequency: How often do students report working with same- and other-gender peers? (RQ1). Based on the tendencies of students to prefer same-gender interactions and to believe that they share similarities with same-gender peers more so than with other-gender peers (Martin et al., 2011), we expected that students would report more often working with same-gender than other-gender classmates (Hypothesis 1). With gender-segregated play outside the classroom increasing over middle childhood (Maccoby & Jacklin, 1987), we thought it possible that students would report increases in gender-segregated interactions over the academic year (Hypothesis 2) and would be even more likely to increase across grades (Hypothesis 3).

The second research question we addressed concerned interaction quality: How well do students report working with same- and other-gender classmates? (RQ2). Given the interaction history that most children have of spending more time with same-gender peers than with other-gender peers outside of school and research showing that children report feeling more comfortable with same- than with other-gender peers (Zosuls et al., 2014; Zosuls et al., 2011), we expected a gender-segregated pattern. Specifically, we expected that children are likely to

report feeling that they work better with same- than with other-gender peers (Hypothesis 4), and, because of these early positive experiences, we might expect that this positivity toward same-gender peers relative to other-gender peers will increase over time (Hypothesis 5) and grade (Hypothesis 6).

We also explored the interrelations over time between interaction frequency and quality, assessing both directions of prediction. Such descriptive analyses may provide insights into potential interventions by assessing whether interactions with peers early in the school year relates to having more positive experiences with them later in the year, and the reverse, whether having high-quality interactions early in the school year relates to increased interactions later in the year. Specifically, we examined whether interaction frequency in the fall predicted interaction quality in the spring for both same- and other-gender peer interactions and whether quality of interaction in the fall predicted interaction frequency in the spring for both same- and other-gender peer interactions (RQ 3). Bidirectional patterns also might be apparent. According to intergroup contact theory (Allport, 1954; Pettigrew, 1998; Pettigrew et al., 2011), early experiences with other-gender peers should lead to improved attitudes, which likely would influence the quality of interactions later in the year. Quality of interactions also improves attitudes; in a study of children's other-gender friendships, bidirectional effects between attitudes and interaction were identified (Citation withheld for blind review).

Method

Participants

We first obtained permission of school districts and then were given a list of potential schools. From those schools, we selected three that adhered most closely to the population demographics of the state and contacted principals to discuss the study. Minority enrollment in

these schools was 56-69% (state average 61%) (see below) and 32-54% of students were eligible for reduced lunch (state average 33%) and 7-11% for free lunch (state average 11%). Once approved, we contacted teachers and then parents to discuss the study and to provide consent letters.

Participants were 515 students in 3rd (174, 33.8%), 4th (177, 34.4%), and 5th (164, 31.8%) grade (*Mage* = 9.08, *SD* = 1.00). Parents reported on child's gender and so did students themselves. Just over half were reported by parents to be boys (51.1%; 1 of them was identified as a socially transitioning child, [trans-boy]); the rest were reported by parents to be girls. The proportion of girls in class ranged from 37% to 67% (average proportion was 49%). Participants were drawn from 26 classrooms in three public elementary schools in Southwest U.S., with 7, 10, and 12 classrooms in each school. The average number of participating students per classroom was 21 and the average classroom participation rate was 70.3%, ranging from 40% to 92.4%. Based on parent report of their child's ethnicity, there were 242 (47%) non-Latinx students, 221 (42.9%) Latinx students, and the remaining 52 (10.1%) students had unknown ethnicity. As for race, there were 243 (47.2%) White students, 72 (14%) Black students, 45 (8.7%) American Indian or Alaska Native students, 9 (1.7%) Asian students, 2 (.4%) Native Hawaiian or Pacific Islander students, 51 (9.9%) multiracial students, 2 (.4%) students were "other", and 91 (17.7%) had unknown race based on parent report.

Attrition. A total of 477 students had data at both time points, and 27 students attrited at T2. Pearson Chi-square tests indicated that attrited students did not differ from non-attrited students with regard to child gender, grade, ethnicity (coded as Latinx, non-Latinx, versus Other/Unknown). Independent samples t-test showed that attrited students did not differ from non-attrited students in terms of same- and other-gender interactional quality and frequency.

Procedures

The study procedures were approved by participating school districts and the Institutional Review Board (IRB) of the university. We used a short-term longitudinal design and collected data in October/Fall (T1) and May/Spring (T2) over the course of one academic year. The data were collected in the year before the COVID-19 pandemic began. Parental consent for children's participation was obtained and children assented at the beginning of data collection. Students filled out a paper-and-pencil survey in a large group setting. Each student had a privacy board and students from different classes were mixed in their seating to further increase privacy. One research staff read the questionnaires aloud while several other trained research assistants answered students' questions and supported students with reading/writing difficulties. Students received a small gift after each assessment.

Measures

Interaction Frequency and Quality

At each assessment, students reported how often they work with same- and other-gender classmates during classroom activities over the past month and how well they interact with other- and same-gender classmates on classroom activities the past month using the Academic Peer Interactions Rating Scale (Singleton & Asher, 1977). Specifically, students were given a class roster and asked to rate 2 items for each classmate on a 5-point scale. The items were (1) In the last month, how often did you work with NAME on classroom activities (*Frequency;* 0 = *never,* 1 = *one time,* 2 = *at least once a week,* 3 = *a few times a week,* 4 = *a lot*); (2) In the last month, how much did you enjoy working with NAME on classroom activities (*Quality;* 0 = *never* to 4 = *a lot*). Composites of interaction frequency and quality were calculated by averaging each student's reports of interactions with same- and other-gender classmates. Further, difference

scores of same-gender minus other-gender classmates (for both frequency and quality) were calculated. The methods we used were similar to sociometric approaches, which are found to be reliable and valid indicators of students' experiences (Parker & Asher, 1993), and has been used in research with late elementary students (Citation withheld for blind review). Although traditional methods of calculating reliability could not be used here due to one-item responses, the items assessed in the fall and spring moderately correlated with one another, indicating test-retest reliability (*rs* =.40 - .53) for these measures. Furthermore, the validity of student reports has been supported: student reports of class groupings have been shown to correspond well with teacher reports and with observations of interactions, suggesting students are able to make these determinations (Gest, 2006; Gest et al., 2003).

To further assess validity, we examined convergent validity by correlating interactional quality and students' school liking, which consisted of 10 items and has demonstrated good psychometric properties (Ladd et al., 2000; Ladd & Price, 1987). We expected that children who engaged in high quality interactions with peers would also like school more. We found that interaction quality was significantly and positively related to school liking as expected, $r(465) = .25$ at T1, $p < .001$ and $r(475) = .32$, $p < .001$ at T2, providing evidence for convergent validity.

Analytic Plan

Except where specified, analyses were conducted with SPSS version 27. We first examined basic descriptive information including the means and correlations among variables. For the classroom analyses, given that these were exploratory analyses, we applied data visualization methods to examine both the variability and levels of gender-segregated patterns in classroom-averaged measures.

To test hypotheses, we first calculated intraclass correlations (ICCs) for main study variables (i.e., students' interaction quality and frequency). ICCs were 8.97% for T1 same-gender frequency, 7.27% for T2 same-gender frequency, 3.72% for T1 same-gender quality, 1.08% for T2 same-gender quality, 6.77% for T1 other-gender frequency, 8.21% for T2 other-gender frequency, 11.90% for T1 other-gender quality, and 9.01% for T2 other-gender quality. According to Hox (1998), ICCs above 5% (.05) indicate data interdependency. Thus, we accounted for class-level variance by including classroom as a covariate in subsequent mixed-model ANCOVAs.

To test whether gender segregation was evident in either frequency of interactions (Hypothesis 1-3) or quality of interactions (Hypothesis 4-6), we performed a repeated measures analysis of variance with two between-subjects factors: Participant Gender (girls, boys) and Grade (3^{rd}, 4^{th}, 5^{th} grade), and two within-subject factors: Type (gender of peer: same-gender, other-gender) and Assessment Time (Time1, fall, Time2, spring), with classroom as a covariate. When interactions were significant, we probed by conducting simple effects analyses. Further, to assess the degree of segregation, we conducted two repeated measures ANCOVAs to assess how the difference scores between same- and other-gender frequency and quality, respectively, changed by time, gender, and grade (controlling for classroom).

To explore how interaction quality influenced later interaction frequency and how frequency of interaction related to later interaction quality, path analyses (using Mplus) were conducted to explore how each measure in the fall predicted the other in the spring.

Results

Descriptive Analyses

Means and Correlations for Students' Ratings

Means and standard deviations for students' interaction frequency and quality at T1 and T2 are presented in Table 1 for the total sample, and by child gender, and child grade. There is a consistent trend of same-gender scores being higher than other-gender scores both for interaction quality and interaction frequency. Table 2 contains the correlations among main variables, separately for girls and boys. Interaction quality and frequency were positively correlated with each other, within and across time for both girls and boys, although there were some minor variations.

Classroom Variability

To understand the extent of gender segregation, we examined the variation among classes on both frequency and quality of interactions among same- and other-gender classmates. The goal was to assess how many classes showed an overall pattern of gender segregation, using class averages for the variables of interest. As can be seen in Figure 1 and Figure 2 (in which classes are ordered by degree of same-gender frequency or quality), although there is considerable variability in the degree of gender segregation, every class (across the 26 classrooms) showed higher same-gender interaction and quality than other-gender interaction and quality. Notably, two classes of the 26 (8%) showed much less gender segregation than the rest of the classes.

Hypothesis Testing: Child-Level Interaction Frequency (RQ1)

Hypothesis 1. At the child level, we predicted we would find that children would evidence gender segregated interactions, in this case, higher levels of interaction with same- than with other-gender classmates (Hypothesis 1). In support of our hypothesis, we found that same-gender interactions were reported to occur more often than other-gender interactions, $F(1, 451) = 5.50$, $p < .05$, $\eta p^2 = .01$ (small effect size; see Figure 3). This effect was qualified by several

significant interactions, although none changed the overall pattern of same-gender interactions being more frequent than other-gender interactions. The only gender difference was for 4th graders, in which girls had greater same-gender interaction frequency than did boys ($p < .05$), and no difference was found for other-gender interactions. For all other grades, girls and boys did not differ in frequency of same- versus other-gender peer interactions.

Hypothesis 2. We expected that gender-segregated interactions would become greater over time (e.g., increasing frequency of same-gender interactions, decreasing frequency of other-gender interactions), which would be evident in a Time X Type interaction (Hypothesis 2). We did not find this interaction to be significant but we did find a significant Time main effect: frequency of interaction increased over Time (fall to spring) for both same- and other-gender interactions, $F(1, 451) = 4.02$, $p = .045$, $\eta p^2 = .01$ (small effect size), but interaction patterns did not vary depending on whether they were same- or other-gender peers. As such, there was no evidence that gender segregation changed over time; instead, it remained relatively stable across the short-term longitudinal design of this study. Thus, Hypothesis 2 was not supported.

Hypothesis 3. We also expected that older students in higher grades would show greater gender segregation (Hypothesis 3), which would be evident by a significant Grade X Type interaction with the difference between same-gender interaction frequency and other-gender interaction frequency being larger with higher grades. Hypothesis 3 was partially supported because we found the expected pattern for girls but not boys. Probing the significant three-way interaction (i.e., Type X Gender X Grade) showed that Grade X Type was significant for girls, $F(2, 224) = 13.29$, $p < .001$, $\eta p^2 = .11$ (large effect size), but not for boys, $F(2, 226) = 1.43$, $p = .242$, $\eta p^2 = .01$ (small effect size). The pattern for girls only partly conformed to Hypothesis 3: 4th grade girls rated same-gender interaction frequency higher than did girls in 3rd ($p < .05$), but

contrary to expectations, 4th grader girls also rated same-gender frequency higher than did 5th grade girls (p <= .001) (3rd and 5th girls did not differ). We found no differences between grades on other-gender interaction frequency (ps > .05). So, minimal support was found with an increase in gender segregation by grade though it was only for girls and only from 3rd to 4th grade.

Hypothesis Testing: Child-Level Interaction Quality (RQ2)

Hypothesis 4. At the child level, we expected to find gender-segregated patterns in quality of interactions such that children's same-gender interactions would be perceived as being more positive than other-gender interactions (Hypothesis 4). We found support for this pattern in the significant Type main effect, $F(1, 447) = 20.07$, $p < .001$, $\eta p^2 = .04$ (moderate effect size). There was a significant Type X Gender interaction effect, $F(1, 447) = 44.72$, $p < .001$, $\eta p^2 = .09$ (moderate effect size), and the other two-way, three-way, and four-way interactions were not significant.

Because of the significant interaction with gender, we then examined whether both boys and girls showed the pattern of same-gender interactions being more positive in quality than other-gender interactions. When probed, the Type effect was significant for each gender (ps < .001) but girls rated same-gender (not other-gender) interactions as more positive than did boys ($p < .001$). When gender differences within each type were examined, gender differences were found only for the quality of same-gender interactions. For these interactions, girls rated them as more positive than did boys ($p < .001$), and there was no gender difference in the quality of other-gender interactions ($p = .259$). Figure 4 presents means by gender and grade.

Hypotheses 5 and 6. We found no support that gender segregation in quality of interactions would become stronger over time (Hypothesis 5) as there was no Time main effect,

$F(1, 447) = .08$, $p = .774$, $\eta p^2 = .000$ (small effect size), or any significant interactions with Time. Further, there were no significant interactions with Grade (Hypothesis 6).

Examining Interrelations of Frequency and Quality Over Time (RQ3)

To explore the direction of effects over time between frequency and quality, we conducted a path analysis in *M*plus version 8.4. Specifically, we examined whether interaction frequency in the fall predicted interaction quality in the spring for both same- and other-gender peer interactions and whether quality of interaction in the fall predicted interaction frequency in the spring for both same- and other-gender peer interactions (RQ 3). This path model was fully saturated. As can be seen in Figure 5, interaction frequency in the fall (with same-or other-gender peers) did not predict how well same- or other-gender students worked together in the spring (*ps* ranged from .07 to .94). Overall, students' interaction frequency did not show a strong relation to their later interaction quality with peers. However, we found evidence for the reverse pattern in that higher quality interactions with other-gender peers in the fall related to later more frequent other-gender interactions in the spring ($p < .001$), whereas having higher quality interactions with same-gender peers in the fall *negatively* related to frequency of other-gender peer interactions in the spring ($p < .009$).

Discussion

In the present study, our goal was to revisit an old but still highly relevant issue in elementary school education—gender segregation (GS)—which is often neglected and yet consequential. Classic studies conducted decades ago, in which children's friends in school and preferred classmates to work with were of the same-gender in most learning groups, suggested that GS was commonly found (e.g., Tuma & Hallinan, 1979). We asked whether that is still the case today.

Our major hypotheses centered on the expectation that GS would occur in contemporary elementary school classrooms as evidenced by both classroom and individual level patterns of interaction frequency and interaction quality and that levels of GS would increase over the school year. We found general support for these hypotheses: Gender segregation occurred consistently across grade, gender, and time although the effects were small for frequency of interactions and moderate for quality of interactions. Further, earlier higher quality other-gender interactions enhanced the possibility of later increased frequency of other-gender interactions but the reverse was not found. We discussed the findings in more depth below.

Patterns of Gender Segregation Indicated by Frequency of Interaction

Probably the clearest and most intuitive measure of GS is *how often* same- and other-gender students interact and work with each other in their classes. At the child level, for every grade, for each gender, and over both fall and spring, a strikingly consistent interaction pattern was found with students reporting that same-gender interactions occurred significantly more frequently than other-gender interactions. The only variation in this pattern was that same-gender interactions were somewhat more common for 4th grade girls than for other children, which partially supported the hypothesis that older students would show stronger GS than younger students. At the classroom level, every classroom showed GS, although the extent of difference between same- and other- interactions varied considerably.

A question for future research is to better understand the correlates of this tendency toward gender segregation. Is this pattern due to teachers, to students, or to both? Do some teachers actively try to offset children's tendencies to gender segregate by structuring their classrooms by mixed-gender groups whereas other teachers pay less attention to this issue, or is this a feature of the children who happen to belong to particular classes? Many studies have

documented the general tendency for children and adolescents in out-of-school contexts to show preferences for same-gender interactions (Mehta & Strough, 2009, 2010) and these out-of-school preferences may be brought into the classroom and be responded to by teachers. Studies of both teachers' use of segregated versus integrated seating and grouping strategies and of students' preferences for working partners within classrooms may better illuminate the causes of classroom variability.

Regarding changes over time in GS, contrary to our hypothesis that same-gender interactions would increase and other-gender interactions decrease, all interactions increased in frequency from fall to spring. It may be that once established, these interaction patterns become comfortable and therefore are repeated. However, it is also important to keep in mind that the overall levels of peer interaction were low for both same- and other-gender interactions, and this has to be taken as the backdrop for interpreting the present findings.

The low level of interaction students reported speaks to another issue important in education, that is, the extent to which group-learning situations are used in elementary classrooms. The findings are surprisingly similar to those found in the classic studies of GS conducted 40 years ago. For example, similar to Lockheed and Harris's (1982) finding that students had less than one mixed-gender (GI) interactions, in the present study, the mean interaction frequency reported by students was 1 to 1.5, which falls between "one time" to "at least once a week" in the previous month. Not only are there limited opportunities for students to interact with other-gender peers in their working groups, our data suggest that there are also limited opportunities for students to work with other students in general regardless of gender. That is, in our sample, group learning was not reported to be used extensively as relatively low levels of interaction were reported for both same- and other-gender interactions. Students in

these classes appeared to be engaged in independent learning and reported having few opportunities for group learning. Why this is the case is unclear. The present data were collected from a representative group of classes prior to the COVID-19 pandemic and yet show that working groups in these elementary school classrooms were not very common, at least as reported by the children themselves. It may be that students fail to register the situations in which they work with their classmates, or it may be that teachers were not frequently promoting this type of learning in their classrooms. With the rise in class size commonly occurring, group-based learning situations might be harder to manage than class-based approaches. This is unfortunate given that research has demonstrated that such group cooperative learning is advantageous for students in the short- and long-term by improving students' metacognition, memory, and strategy use (Manion & Alexander, 1997, 2001). Furthermore, with the changes in educational settings due to the pandemic and increased need for safety and social distancing, students may have even fewer opportunities in the future to engage in these cooperative learning groups and reduced opportunities for gender integration experiences.

Quality of Interaction as an Indicator of Gender Segregation

Another measure of GS included in the present study was the *quality* of same- and other-gender interactions. As hypothesized, students showed gender-segregation in that they reported higher quality interactions with same- versus other-gender classmates with a moderate effect size, and these did not vary by grade or time. A gender difference was found, however, with girls reporting more same-gender positivity than did boys. This gender difference may reflect that girls hold stronger trait-based gender stereotypes than boys, possibly colored by a generalized "boys are bad" stereotype (Heyman, 2001). Furthermore, girls may learn through experience that they are less able to influence the outcomes of group work when they work with other-gender

classmates (Lockheed & Harris, 1984) and this may decrease their perceptions of having positive interactions with boys. Although mixed-gender group work might be more challenging for girls if they perceive fewer opportunities for leadership or influence, when they do have positive experiences with and effectively problem solve in these circumstances, it may be especially beneficial for girls (DiDonato et al., 2016).

Corresponding to students' individual reports were the classroom-level data that also showed virtually every class reporting higher quality of same-gender peer interactions when working together compared to when working with other-gender peers (see Figure 2). Only two classes showed similar quality ratings for same- and other-gender peers.

The student-reported and classroom-level findings concerning the quality of interactions with other students may depend in part on the structure of learning in the classroom. Cooperative working groups provide support and structure for students as compared to individual work, and it is in these types of learning situations that mixed-gender interactions may be the most positive (Slavin, 1995). And even within cooperative groups, ensuring that students have training in how to interact effectively in small groups (e.g., listening skills, providing constructive feedback) sets the stage for increased student success in mixed-gender groups as compared to placing students together without such training (Gillies & Ashman, 1998). These same skills have been found useful for promoting more diverse peer interactions and improving academic performance in 5th-grade students using a year-long relationship skills training intervention (DeLay et al., 2016). To the extent that teachers employ this social-emotional type of learning of relationship skills, we would expect to find that quality of all group interactions improves.

Interrelations of Indicators of Gender Segregation Over Time

Understanding how the frequency and quality of interactions with same- and other-

gender peers relate to one another over time may provide helpful insights into educational practices. Based on intergroup contact theory and research (Pettigrew, 1998; Pettigrew & Tropp, 2006), we expected that increased contact with other-gender peers (an outgroup) would lead to improved quality of those relationships over time, however, this pattern was not supported. Nonetheless, these findings are consistent with earlier work indicating that infrequent mixed-gender groupings were not effective in improving students' attitudes about mixed-gender work (Lockheed & Harris, 1984). It is possible that a high level of involvement and support from teachers as well as consistent, long-term, and frequent use of mixed-gender groupings are required to ensure the effectiveness of these groups (see Lockheed & Harris, 1982). With less support and less frequent practice, there may not be sufficient opportunities to promote more positive stereotypes and attitudes about other-gender peers. For example, recent research suggests that when teachers support frequent mixed-gender groupings or pairings of students in class, students benefit from those interactions by increasing the diversity of their interaction partners (Hanish et al., 2021).

The present results demonstrated that when students reported higher quality interactions with other-gender students in the fall, they later reported more frequent other-gender interactions in the spring. If other-gender students are able to work together well, teachers may be more inclined to form mixed-gender groupings, thereby expanding students' access to these opportunities. In contrast, if these groupings are not seen as working well together, teachers may be less inclined to ask students to work in mixed-gender groups. In fact, some teachers use gender as a punishment in their classes (e.g., moving a rowdy student to sit next to an other-gender peer). Why the same pattern was not seen for same-gender working groups is surprising. Whether teachers are more attentive to or aware of the quality of other-gender as compared to

same-gender interactions and which specific qualities they attend to are important topics for future investigation.

Gender Segregation Revisited: Implications and Why it Matters

GS is a powerful force in our social lives. Individuals of all ages tend to prefer spending time with others they believe are like themselves (Farver, 1996; McPherson et al., 2001; Rubin et al., 1994). GS reinforces gender-typical behaviors (Lockheed & Harris, 1982; Martin & Fabes, 2001), but it also provides children with information about the interaction styles, communication, and interests of their "ingroup." How much time is spent with same-gender peers (or social dosage) matters, however. Ideally, same-gender and other-gender interactions would be more balanced so that children develop a more expanded and inclusive experience of others. As reported in Lockheed (1986, p. 627), one teacher explained this idea to her students and it made a difference. She said, "that "life" required one to work with lots of different types of people and that she would expect her students to "sit like life" in the classroom: boys and girls sitting with and next to each other." This strategy worked for her students: "the children readily accepted this new norm and grouped themselves accordingly". As this teacher realized, children (and adults) need to supplement their ingroup experiences with experiences that differ, and in this case, with other-gender classmates (Halim et al., 2021; Leaper, 1994; Maccoby, 1998). Some educators may think gender has been managed in their classrooms as children are encouraged to cross gender stereotypic boundaries in their activities, but the present research clearly demonstrates that gender continues to be a central organizing feature of peer interactions within classrooms. Returning attention to the importance of this type of diversity exposure for students is essential for educators.

Given the demands to focus on academics, teachers may have limited energy and time to

devote to thinking about peer relationships in their classes. And even if they do attend to these issues, gender may not be the social grouping that is of concern. Once recognizing that gender segregation is occurring, what can teachers do to modify these patterns? Teachers are quite good at shaping the relationships in their classes (Gest & Rodkin, 2011; Kindermann, 2011) and thus are able to modify GI their classrooms. In a few studies, evidence has accrued that teachers have the ability to promote GI by reinforcing mixed-gender interactions when they occur (Lockheed, 1986; Serbin et al., 1977), through teaching relationship skills to their students so they can work more effectively in groups (DeLay et al., 2016), and through structuring and encouraging mixed-gender interactions formally through "buddy" or work groups (Hanish et al., 2021). Students have the potential to benefit in many ways from these efforts: they gain larger social networks, increased academic support (Caprara et al., 2000), and increased understanding of their classmates if they are encouraged to work with a broader array of students.

The present study also suggests that educators and researchers may find that monitoring of students' perceptions of learning opportunities with each member of the class is effective for gaining nuanced insights into gendered peer relationships in classrooms. By tracking students' reports of frequency and the quality of working with same- and other-gender classmates, this would allow a more focused response to the call to attend to GS tendencies and to the promotion of GI. For instance, if students' interactions were monitored while teachers employed various methods of guiding peer relationships in their classes, insights would be gained about the nature of changes that occur. More fine-grained methods for assessing GS and GI will be important for furthering this area of research.

Limitations and Conclusions

In the present study, we examined gender segregation indicators in multiple classrooms

across multiple grades in one city in the United States. Conducting larger-scale studies with more participants, more diversity of participants, more classrooms, over longer time periods, and more varied locations will increase the generalizability of the results. Furthermore, many potential indicators of GS could be adopted; here, we used two, but others are reasonable and may show stronger or weaker levels of GS. The measures we used required retrospective memory for students' interactions; these reports are likely influenced by a variety of factors which we were unable to account for in the present study design. For instance, students' reports might be biased in favor of gendered expectations for interactions or by recent experiences that occur outside the classroom. To offset these concerns, including additional measures of gender integration would be useful such as asking more general questions about students' feelings of inclusion. Ideally, researchers may collect observational data on the gender composition of student working groups, however, these methods are expensive to conduct and may themselves be limited. In particular, observational data may not capture how well students work together. Although the reports we used were retrospective, they did allow for students to have differing experiences with each student in class, and these were relatively stable over the year. As to additional future research efforts, conducting studies with large enough samples to explore the roles of intersecting identities (gender, ethnicity, race) will provide important next steps to fully understanding how students' identities relate to their within-classroom interactions.

Although there was considerable variability across classrooms and there was a limited degree of "classwork" interaction reported by students, in every classroom, we found consistent evidence of GS. Despite societal changes related to gender over the almost 40 years since the earlier studies of GS were published, elementary schools are still marked by GS. It appears that these coeducational classes are not providing a truly coeducational experience for students due to

limited interaction opportunities and due to limited other-gender peer interactions. We hope that this research will alert educators to this situation and begin to change it by promoting more cooperative learning groups and more GI in these groups. As Lockheed and Harris (1984, p. 276) noted then (which still applies today):

> *One plausible explanation for the absence of research on cross-sex cooperation is that unlike racial segregation, for example, sex segregation in elementary schools often is not considered a problem…*

This "problem" of non-coeducational experiences within coeducational classrooms has a chance to be resolved as educators become more knowledgeable about the value of gender integration and about the methods that are likely to be effective in promoting students to "sit like life".

References

Allport, G. W. (1954). *The nature of prejudice.* Addison-Wesley.

Brown, C. S. (2013). Legal issues surrounding single-sex schools in the U.S.: Trends, court cases, and conflicting laws. *Sex Roles, 69*, 356-362. https://doi.org/10.1007/s11199-011-0001-x

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*(4), 302-306. https://doi.org/https://doi.org/10.1111/1467-9280.00260

DeLay, D., Zhang, L., Hanish, L. D., Miller, C. F., Fabes, R. A., Martin, C. L., Updegraff, K., & Kochel, K. P. (2016). Peer influence on academic performance: A social network analysis of social-emotional intervention effects. . *Prevention Science, 17*, 903-913. https://doi.org/10.1007/s11121-016-0678-8

DiDonato, M., Martin, C. L., Hanish, L., Fabes, R. A., & Iida, M. (2016). *Mixed-sex interaction during a science task: Working with boys improves girls' science-related perceptions.* [manuscript in progress]. Arizona State University

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. Public Opinion Pools from 1946 to 2018. *American Psychologist, 75*(3), 301-315. https://doi.org/https://doi.org/10.1037/amp0000494

Fabes, R. A., Martin, C. L., Hanish, L., & DeLay, D. (2018). Gender integration in coeducational classrooms: Advancing educational research and practice. *School Psychology Quarterly, 33*(2), 182-190. https://doi.org/http://dx.doi.org/10.1037/spq0000266

Fabes, R. A., Shepard, A. S., Guthrie, I. K., & Martin, C. L. (1997). Roles of tempermental arousal and gender segregated play in young children's social adjustment. *Developmental Psychology, 33*(4), 693-702. https://doi.org/https://doi.org/10.1037/0012-1649.33.4.693

Farver, J. M. (1996). Aggressive behavior in preschoolers' social networks: Do birds of a feather flock together? *Early Childhood Research Quarterly, 11*, 333-350. https://doi.org/https://doi.org/10.1016/s0885-2006(96)90011-3

Gest, S. D. (2006). Teacher reports of children's friendships and social groups: Agreement with peer reports and implications for studying peer similarity. *Social Development, 15*(2), 248-259. https://doi.org/https://doi.org/10.1111/j.1467-9507.2006.00339.x

Gest, S. D., Farmer, T. W., Cairns, B. D., & Xie, H. (2003). Identifying children's peer social networks in school classrooms: links between peer reports and observed interactions. *Social Development, 12*(4), 513-529.

Gest, S. D., & Rodkin, P. C. (2011). Teaching practices and elementary classroom peer ecologies. *Journal of Applied Developmental Psychology, 32*, 288-296. https://doi.org/https://doi.org/10.1016/j.appdev.2011.02.004

Gillies, R. M., & Ashman, A. F. (1998). Behavior and interactions of children in cooperative groups in lower and middle elementary grades. *Journal of Educational Psychology, 90*(4), 746-757. https://doi.org/http://0022-0663/98

Glazier, J. J., Gulgoz, S., & Olson, K. R. (2020). Gender encoding in gender diverse and gender conforming children. *Child Development, 91*(6), 1877-1885. https://doi.org/10.1111/cdev.13399

Halim, M. L., Martin, C. L., Andrews, N. C. Z., Zosuls, K. M., & Ruble, D. N. (2021). Enjoying each other's company: Gaining other-gender friendshps promotes positive gender attitudes among ethnically diverse children. *Personality and Social Psychology Bulletin, 47*(12), 1635-1653. https://doi.org/https://doi.org/10.1177/0146167220984407https:/

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J. S., Liben, L. S., & Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science, 333*, 1706-1707. https://doi.org/10.1126/science.1205031

Hanish, L. D., Martin, C. L., Cook, R., DeLay, D., Lecheile, B., Fabes, R. A., Goble, P., & Bryce, C. I. (2021). Building integrated peer relationships in preschool classrooms: The potential of buddies. . *Journal of Applied Developmental Psychology, 73*(Mar 2021-Apr 2021). https://doi.org/doi: 10.1016/j.appdev.2021.101257

Heyman, G. D. (2001). Children's interpretation of ambiguous behavior: Evidence for a "boys are bad" bias. *Social Development, 10*, 230-247. https://doi.org/https://doi.org/10.1111/1467-9507.00161

Hox, J. (1998). Multilevel modeling: When and why. In *Classification, data analysis, and data highways* (pp. 147-154). Springer. https://doi.org/https://doi.org/10.1007/978-3-642-72087-1_17

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2018). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171-193. https://doi.org/http://dx.doi.org/10.1037/amp0000307

Kindermann, T. A. (2011). Commentary: The invisible hand of the teacher. *Journal of Applied Developmental Psychology, 32*, 304-308. https://doi.org/10.1016/j.appdev.2011.04.005

La Freniere, P., Strayer, F. F., & Gauthier, R. (1984). The emergence of same-sex affiliative preferences among preschool peers: A developmental/ethological perspective. *Child Development, 55*, 1958-1965. https://doi.org/https://doi.org/10.2307/1129942

Leaper, C. (1994). Exploring the consequences of gender segregation on social relationships. In C. Leaper (Ed.), *Childhood gender segregation: causes and consequences* (pp. 67-86). Jossey-Bass.

Lockheed, M. E. (1986). Reshaping the social order: The case of gender segregation. *Sex Roles, 14*(11-12), 617-628. https://doi.org/https://doi.org/10.1007/BF00287692

Lockheed, M. E., & Harris, A. M. (1982). Classroom interaction and opportunities for cross-sex peer learning in science. *Journal of Early Adolescence, 2*(2), 135-143. https://doi.org/https://doi.org/10.1177/027243168200200205

Lockheed, M. E., & Harris, A. M. (1984). Cross-sex collaborative learning in elementary classrooms. *American Educational Research Journal, 21*(2), 275-294. https://doi.org/https://doi.org/10.3102/00028312021002275

Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Belknap Press.

Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. In W. R. Hayne (Ed.), *Advances in child development and behavior, Vol. 20.* (pp. 239-287). Academic Press, Inc.

Manion, V., & Alexander, J. M. (1997). The benefits of peer collaboration on strategy use, metacognitive causal attribution, and recall. *Journal of Experimental Child Psychology, 67*, 268-289. https://doi.org/0022-0965/97

Manion, V., & Alexander, J. M. (2001). The benefits of peer collaboration: A replication with a delayed posttest. *Contemporary Educational Psychology, 26*, 588-601. https://doi.org/doi:10.1006/ceps.2001.1082

Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology, 25*(1), 80-88. https://doi.org/ https://doi.org/10.1037/0012-1649.25.1.80

Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology, 37*, 431-446. https://doi.org/10.1037/0012-1649.37.3.431

Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: children's beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships, 16*, 751-771. https://doi.org/10.1177/0265407599166005

Martin, C. L., Fabes, R. A., & Hanish, L. D. (2014). Gendered-peer relationships in educational contexts. In L. S. Liben & R. S. Bigler (Eds.), *Advances in Child Development and Behavior* (Vol. 47, pp. 151-187). Elservier.

Martin, C. L., Fabes, R. A., Hanish, L. D., Leonard, S., & Dinella, L. (2011). Experienced and expected similarity to same-gender peers: Moving toward a comprehensive model of gender segregation. *Sex Roles, 65*, 826-842. https://doi.org/10.1007/s11199-011-0029-y

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444. https://doi.org/https://doi.org/10.1146/annurev.soc.27.1.415

Mehta, C. M., & Strough, J. (2009). Sex segregation in friendships and normative contexts across the life span. *Developmental Review, 29*(3), 201-220. https://doi.org/https://doi.org/10.1016/j.dr.2009.06.001

Mehta, C. M., & Strough, J. (2010). Gender segregation and gender-typing in adolescence. *Sex Roles, 63*(3-4), 251-263. https://doi.org/https://doi.org/10.1007/s11199-010-9780-8

Miller, C. F., & Updegraff, K. (2014). *Teachers' views of gender in their classes.* [unpublished data]. Arizona State University.

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*, 611-621. https://doi.org/https://doi.org/10.1037/0012-1649.29.4.611

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65-85. https://doi.org/https://doi.org/10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751-783. https://doi.org/https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations, 35*, 271-280. https://doi.org/10.1016/j.ijintrel.2011.03.001

Rubin, K. H., Lynch, D., Coplan, R., Rose-Krasnor, L., & Booth, C. L. (1994). "Birds of a feather...": Behavioral concordances and preferential personal attraction in children. *Child Development, 65*(6), 1778-1785. https://doi.org/https://doi.org/10.2307/1131293

Serbin, L. A., Moller, L. C., Gulko, J., Powlishta, K. K., & Colburne, K. A. (1994). The emergence of gender segregation in toddler playgroups. In C. Leaper (Ed.), *Childhood gender segregation: Causes and consequences. New directions for child development, No. 65.* (pp. 7-17). Jossey-Bass Inc, Publishers, San Francisco, CA, US.

Serbin, L. A., Tonick, I. J., & Sternglanz, S. H. (1977). Shaping cooperative cross-sex play. *Child Development, 48*, 924-929. https://doi.org/https://doi.org/10.2307/1128342

Singleton, L. C., & Asher, S. R. (1977). Peer preferences and social interaction among third-grade children in an integrated school district. *Journal of Educational Psychology, 69*, 330-336. https://doi.org/https://doi.org/10.1037/0022-0663.69.4.330

Slavin, R. E. (1995). Enhancing intergroup relations in schools: Cooperative learning and other strategies. In W. D. Hawley & A. W. Jackson (Eds.), *Toward a common destiny: Improving race and ethnic relations in America*. Jossey-Bass.

Strough, J., Swenson, L. M., & Cheng, S. (2001). Friendship, gender and preadolescents' representation

of peer collaboration. *Merrill-Palmer Quarterly, 47*(4), 475–499. https://doi.org/10.1353/mpq.2001.0025

Tuma, N. B., & Hallinan, M. T. (1979). The effects of sex, race, and achievement on schoolchildren's friendships. *Social Forces, 57*(4), 1265-1285. https://doi.org/https://www.jstor.org/stable/2577270

Webb, N., Baxter, G., & Thompson, L. (1997). Teachers' grouping practices in fifth-grade science classrooms. The Elementary School Journal, 98, 91-113.

Zosuls, K. M., Field, R. D., Martin, C. L., Andrews, N. C. Z., & England, D. E. (2014). Gender-based relationship efficacy: Children's self-perceptions in intergroup contexts. *Child Development, 85*, 1663-1676. https://doi.org/10.1111/cdev.12209

Zosuls, K. M., Martin, C. L., Ruble, D. N., Miller, C., Gaertner, B. M., England, D. E., & Hill, A. P. (2011). "It's not that we hate you": Understanding children's gender attitudes and expectancies about peer relationships. *British Journal of Developmental Psychology, 29*, 288-304. https://doi.org/10.1111/j.2044-835X.2010.02023.x