Controlling for Prior Attainment Reduces the Positive Influence that Single-sex Classroom Initiatives Exert on High School Students' Scholastic Achievements

Charlotte R. Pennington

Lancaster University

Linda K. Kaye, Adam W. Qureshi, and Derek Heim

Edge Hill University

Author Note

Charlotte R. Pennington, Department of Psychology, Lancaster University; Linda K. Kaye, Department of Psychology, Edge Hill University; Adam W. Qureshi, Department of Psychology, Edge Hill University; Derek Heim, Department of Psychology, Edge Hill University

Corresponding concerning this manuscript should be addressed to Charlotte R. Pennington, Department of Psychology, Lancaster University, Bailrigg, Lancashire, United Kingdom, LA4 1YW. E-mail: c.r.pennington@lancaster.ac.uk or charlotte.pennington@outlook.com

Abstract

Research points to the positive impact that gender-segregated schooling and classroom initiatives exert on academic attainment. An evaluation of these studies which reveal positive effects highlights, however, that students are typically selectively assigned to single- or mixed-gender instructional settings, presenting a methodological confound. The current study controls for students' prior attainment to appraise the efficacy of a single-gender classroom initiative implemented in a co-educational high school in the United Kingdom. Secondary data analysis (using archived data) was performed on 266 middle-ability, 11–12 year-old students' standardized test scores in Languages (English, foreign language), STEM-related (Mathematics, Science, Information and Communication Technology), and Non-STEM subjects (art, music, drama). Ninety-eight students (54, 55% female) were taught in single-gender and 168 (69, 41% female) in mixed-gender classrooms. Students undertook identical tests irrespective of classroom type, which were graded in accordance with U.K national curriculum guidelines. Controlling for students' prior attainment, findings indicate that students do not appear to benefit from being taught in single-gender relative to mixed-gender classrooms in Language and STEM-related subjects. Young women benefitted from being taught in mixed-gender relative to single-gender classes for Non-STEM subjects. However, when prior ability is not controlled for, the intervention appears to be effective for all school subjects, highlighting the confounding influence of selective admissions. These findings suggest that gender-segregated classroom initiatives may not bolster students' grades. It is argued that studies that do not control for selection effects may tell us little about the effectiveness of such interventions on scholastic achievement.

Keywords: single-sex classrooms, co-education, achievement, gender-achievement gap, education policy

Controlling for Prior Attainment Reduces the Positive Influence that Single-sex Classroom Initiatives Exert on High School Students' Scholastic Achievements

The gender-achievement gap is well documented in Western cultures across a number of different subject domains (Else-Quest, Hyde, & Linn, 2010; Stoet & Geary, 2013). Current research indicates that females outperform males typically across the majority of school subjects (Mullholland, Hansen, & Kaminski, 2004; Voyer & Voyer, 2014), particularly in English literacy (Young-Suk, Al Otaiba, Wanzek, & Brandy, 2015). However, there is considerable variation when exploring gender differences in mathematics, with females underperforming in comparison to males at the high end of the distribution (Ceci & Williams, 2010; Reilly, Neumann, & Andrews, 2015; Stoet & Geary, 2013; Wai, Cacchio, Putallaz, & Makel, 2010).

Many factors have been proposed to account for differences in females and males' academic performance. For example, boys tend to report higher academic self-efficacy in mathematics (Dai, 2001), whereas girls report higher self-efficacy in English literacy (Pajares, & Valiante, 2001; Niederle & Vesterlund, 2010). Furthermore, children's academic self-efficacy has been found to be correlated with parents' and teachers' beliefs of gender-subject competence (Bleeker & Jacobs, 2004; Miller, Eagley, & Linn, 2015; Tiedemann, 2002; Wood, Kurtz-Costers, Rowley, & Okeke-Adeyanju, 2010). Gender differences in academic attainment may arise due to the format of achievement tests, with research suggesting that boys excel on standardized tests relative to girls who do better in coursework-based examinations (Ceci, Williams, & Barnett, 2009; Kimball, 1989). In attempt to reduce performance clefts, other research has moved beyond these factors to examine the direct role of the learning context (Park, Behrman, & Choi, 2013; Sullivan, Joshi, & Leonard, 2010). Despite being met with considerable controversy (Bigler & Signorella, 2011; Pahlke et al.,

2013; 2014; Signorella & Bigler, 2013), one solution that has been proposed is single-sex schooling.

Proponents of single-sex schooling suggest that the segregation of females and males has a positive impact on their academic self-concept (Sullivan, 2009), educational transition (Lee & Marks, 1990; Park et al., 2013), and attainment and interest (Else-Quest & Peterca, 2015). Other research indicates that females benefit more from single-sex schooling compared to males (Alon & Gelbgiser, 2011; Else-Quest & Peterca, 2015; Lee & Bryk, 1986; Mullholland et al., 2004), with such environments suggested to lessen the impact of gender stereotypes on females' interest and performance in STEM-related subjects (Inzlicht & Ben-Zeev, 2003; Shapka & Keating, 2003). For example, females report higher competence beliefs and tend to achieve higher grades in mathematics and science when they are taught in single-sex relative to co-educational schools (Eisenkopf, Hessami, Fischbacher, & Ursprung, 2015; Hoffman, 2002).

Those taught in single-sex schools also report fewer experiences of gender stereotyping compared to their mixed-sex counterparts (Pahlke et al., 2014). Gender-segregated learning environments have therefore been suggested to alleviate experiences of *stereotype threat*, a situational phenomenon whereby females apprehend that their performance will be evaluated in line with gender-related expectations (Elizaga & Markman, 2008; Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000; 2003; Picho & Stephens, 2012). Some research indicates that gender-segregated education has a neutral impact on males' academic attainment (Sullivan et al., 2010), whereas other research suggests that males benefit more from being taught in co-educational settings (Jackson & Smith, 2000; Schneider & Coutts, 1982).

Opponents of such educational initiatives, however, argue that single-sex schooling may exacerbate gender stereotyping because students question why they have been separated

from their other-sex peers (Bigler & Liben, 2006; 2007; Halpern et al., 2011). Such environmental cues may, explicitly or implicitly, relay a message to students that gender is a fixed attribute of ability (Dweck, 2008), which has been shown to have a deleterious impact on performance outcomes (Dar-Nimrod & Heine, 2006; Pennington & Heim, 2016). From a developmental perspective, research also suggests that single-sex schooling may come at a longer-term cost to successful gender-role socialization and intergroup cooperation once females and males are eventually re-integrated in ensuing education and workplace settings (Fabes, Lynn, & Martin, 2015; Halpern et al., 2011; Martin & Fabes, 2001).

Due to a number of substantial methodological weaknesses, researchers have argued that studies evaluating the potential efficacy of single-sex schooling need to be interpreted with caution (Halpern et al., 2011; Pahlke et al., 2013; 2014). The most pertinent issue is that many single-sex schools employ selective admissions procedures whereby students are recruited based on their previous ability and socio-economic background (Hayes, Pahlke, & Bigler, 2011; Marsh, 1989; Signorella et al., 2013). However, many studies do not control for selection effects within their analyses (c.f., Pahlke et al., 2014 for a meta-analysis). This greatly undermines the conclusions that can be drawn from research investigating the possible impact that single-sex schooling may have on educational outcomes because students who attend these schools may differ from those attending co-educational schools in important ways (Hayes et al., 2011). Demonstrating the significance of this problem, Pahlke et al. (2014) conducted a meta-analysis and found that studies which did not control for students' previous attainment showed a moderate positive effect of single-sex schooling for mathematics. On the other hand, their findings indicate that studies which controlled for prior achievement tend to show a negligible effect of single-sex classroom settings on attainment levels. They conclude that findings from high quality studies do not support the view that single-sex schooling provides benefits over and above co-educational schooling.

Presenting as an additional issue, research typically compares the effects of the school environment between single-sex and co-educational schools and generalizes these findings across nations (Baker, Riordan, & Schaub, 1995). This creates a number of possible confounds, specifically with regard to the likelihood of differences emerging as a result of variations between school settings and the broader context in which learning takes place (Mael, Alonso, Gibson, Rogers, & Smith, 2005; Shapka, 2009). Consequently, it is difficult to determine whether gains in academic attainment are the result of gender-segregation strategies or the product of other educational variables, such as the social and cultural environment in which students are taught (Pahlke et al., 2014).

On a more practical level, the creation of single-sex schools is influenced heavily by the organization of state education and broader economic factors. For example, the number of single-sex schools in the United Kingdom decreased by approximately 80% in the last three decades of the 20th century because schools received considerable pressure to teach boys and girls jointly to sustain economic viability (Younger & Warrington, 2006). As a consequence, it has been argued that this can make it challenging for teachers to tailor instructional strategies to the presumed different learning needs of females and males in certain subjects (Parker & Rennie, 2002). For example, research indicates that teachers are able to spend more time supporting boy's English performance in single-sex classrooms, as well as manage behavior more effectively (Parker & Rennie, 2002).

The implementation of single-sex classrooms within co-educational schools therefore presents as a potentially viable option to bolster students' participation and performance. Empirical studies appear to show that single-sex classrooms increase females' long-term participation in counter-stereotypical domains such as science and mathematics (Gillibrand, Robinson, Brawn, & Osborn, 1999; Rosenthal, London, Levy, & Lobel, 2011), and bolster males' English proficiency (Parker & Rennie, 2002). However, in their meta-analysis, Pahlke

et al. (2014) failed to find a consistent advantage of single-sex classrooms over single-sex schooling, suggesting that selection effects may confound studies within this area. Given the dearth of research in this area (Arnot, David, & Weiner, 1998; Warrington & Younger, 2003), it is clear that additional research is required to examine reliably the potential effectiveness of single-sex classroom initiatives implemented within co-educational schools.

Building upon this, the current research appraises the efficacy of a single-sex classroom initiative on students' academic attainment in a co-educational high school. This intervention was implemented due to a perceived gender-achievement gap in which teachers reported that girls were outperforming boys in the majority of school subjects. Overcoming the limitations inherent in previous research, the current study controlled for students' prior attainment (pre-intervention), as well as variables relating to socio-economic status, special education needs (SEN), and native language. It was hypothesized that young women would achieve significantly higher grades in Language subjects (Young-Suk et al., 2015), whereas young men would outperform young women in STEM (Stoet & Geary, 2013). Moreover, it was hypothesized that single-sex classrooms would show a positive effect on academic attainment when prior ability was not controlled for, but that these effects would be significantly reduced (if not disappear completely) when accounting for this (Halpern et al., 2011; Pahlke et al., 2013; 2014).

Method

Participants

Data analyses were performed on archived data for 266 students' academic attainment grades, which were obtained throughout their first year of high school (11-12 years of age) in a U.K comprehensive, co-educational school. Of this sample, 123 (46.2%) students were female and 143 (53.8%) were male. A total of 98 students (54 female, 44 male) were placed into single-sex classrooms, with the remaining 168 students (69 female, 99 male) taught in

mixed-sex classrooms. Thirty-six percent of students (n = 96) were registered as having a diagnosis of Special Educational Needs (SEN) (i.e., moderate learning disabilities, attention deficit hyperactivity disorder, dyslexia, autism, hearing impairment, and dyscalculia). Sixty-six received free school meals (FSM; 24.8%), and eight did not speak English as their native language (EAL; 3.0%). None of these factors differed significantly as a function of classroom type or students' gender (all $p > .05$).

Procedure

The school implemented a single-sex classroom initiative with the aim of bolstering students' academic attainment. A letter was sent to the parents of each student explaining the initiative, and parents provided informed consent (through opt-out) for their children to be placed into a single-sex classroom from the start of secondary education. Across the entire sample analyzed, four parents vetoed the procedure and opted for their children to remain in co-educational classroom settings.

The school followed a specific selection criterion to assign students to single-sex or mixed-sex classrooms. Specifically, the school created an average score for each student, using aggregate predicted grades from primary school in English, mathematics and science. They then assigned the highest achieving students (n = 107) to four classrooms of mixed-sex forms. The next 98 students were then placed into four single-sex forms of middle ability, with two all-male and two all-female classrooms. The remaining students were assigned to middle ability, mixed-sex classrooms. Students remained in either single-sex or mixed-sex classrooms for all school subjects, except for Physical Education in which they were taught in single-sex groups. Irrespective of classroom type (single/mixed-sex), students undertook the same standardized tests at the end of the academic year in the subjects of science, mathematics, information and communications technology (ICT), drama, music, English, and foreign language. Students completed an on-going assessment in Art which was graded by

teacher's professional judgement. Although different teachers taught each school subject, the same teachers taught students in both single-sex and mixed-sex classes in their respective subjects.

Analytic Strategy

Given that the school had not assigned randomly students to single-sex or mixed-sex classes, it was important to control for their prior attainment (Pahlke et al., 2014; Pahlke & Hyde, 2016). First, we removed high attaining students (who were all assigned to mixed-sex classrooms, n = 107) from the dataset so that we were left with only middle-attaining students (total n = 266; participant section reports this final number, after exclusions). We then computed a difference score by subtracting students' predicted grades (pre-intervention) from their obtained grades (post-intervention). Students' predicted grades were computed in line with their standardized test scores in primary school and were generated by an external organization. Students' obtained grades represent their standardized test scores in their first year of high school, which were graded in accordance with U.K National Curriculum guidelines (The National Curriculum, 2010). They received a subject-specific attainment level between 1–8, with a higher level indicating better performance. Each of these levels was also split into three ability categories (e.g., Level 4; Lower, Middle, and Upper). For the purpose of statistical analyses, these grades were re-coded from categorical scores to continuous scores on a scale ranging from 1 (Level 2L) to 21 (Level 8U; see Table 1).

An average mean difference was computed for STEM subjects (Science, Math, ICT), non-STEM subjects (Art, Drama, Music), and Languages (English, Foreign Language). This limited the number of analyses conducted and allowed greater control over Type 1 errors compared to analyzing each subject grade separately. Supporting Information File 1 presents analyses for separate school subjects. Data analysis took the form of a 2 (Gender: male, female) x 2 (Classroom type: single-sex, mixed-sex) between-participants Analysis of

Variance (ANOVA). An Analysis of Covariance (ANCOVA) was also conducted to examine whether receiving free school meals (FSM), English as a native language (EAL), and special education needs (SEN) influenced these findings. An adjusted alpha level of $p < .01$ was utilized to elucidate any main effects and interactions. This decision was guided by the rationale that all *p*-values are uniformly distributed under the null hypothesis. As such, an alpha level of $p < .01$ provides stronger evidence against the null hypothesis relative to $p < .05$ and therefore provides more convincing findings (Cumming & Calin-Jageman, 2017, pp. 130). Positive scores indicate that students' obtained grades were higher than their predicted grades, whereas negative scores indicate that their obtained scores were lower than predicted.

Results

Languages

When controlling for prior attainment, there was no significant main effect of classroom type, $F(1, 256) = 1.26$, $p = .263$, $\eta p^2 = .005$, 99% CI [- .12, .31]. There was no significant main effect of gender, $F(1, 256) = .61$, $p = .436$, $\eta p^2 = .002$, 99% CI [- .27, .15]. There was also no significant interaction between gender and classroom type, $F(1, 256) = 4.41$, $p = .037$, $\eta p^2 = .017$. When prior performance was not controlled for, a main effect of classroom type was found, $F(1, 256) = 58.04$, $p < .001$, $\eta p^2 = .19$, with students taught in single-sex classrooms ($M = 9.80$, $SD = 1.42$) appearing to outperform those in mixed-sex classrooms ($M = 7.97$, $SD = 1.98$), $p < .001$, 99% CI [- 2.26, - 1.11]. This highlights the confounding influence of selective admissions. Including FSM, EAL and SEN status as covariates did not significantly influence these findings.

STEM-Subjects (Mathematics, Science, ICT)

Controlling for prior attainment, there was no significant main effect of classroom type, $F(1, 258) = .25$, $p = .617$, $\eta p^2 = .001$, 99% CI [- .24, .36]. There was a significant main effect of gender, $F(1, 258) = 7.31$, $p = .007$, $\eta p^2 = .03$. Simple main effects indicated that male adolescents ($M = -.58$, $SD = .89$) underperformed relative to their predicted grades compared to female adolescents ($M = -.27$, $SD = .88$), $p = .007$, 99% CI [.01, .61]. There was no significant interaction between gender and classroom type, $F(1, 258) = .04$, $p = .850$, $\eta p^2 < .001$. When prior performance was not controlled for, a main effect of classroom type was found, $F(1, 258) = 76.53$, $p < .001$, $\eta p^2 = .23$, with students taught in single-sex ($M = 9.45$, $SD = 1.15$) seemingly outperforming those in mixed-sex classrooms ($M = 7.75$, $SD = 9.45$), $p < .001$, 99% CI [- 2.07, - 1.12]. Including FSM, EAL and SEN status as covariates did not significantly influence these findings.

Non-STEM Subjects (Art, Drama, Music)

There was no significant main effect of classroom type, $F(1, 259) = .058$, $p = .809$, $\eta p^2 < .001$, 99% CI [- .14, .17]. There was no significant main effect of gender, $F(1, 259) = 6.60$, $p = .011$, $\eta p^2 = .025$, 99% CI [- .31, .002]. There was a significant interaction between gender and classroom type, $F(1, 259) = 13.62$, $p < .001$, $\eta p^2 = .05$. Simple main effects indicated that young women underperformed relative to their predicted grades in single-sex ($M = -.30$, $SD = .36$) compared to mixed-sex classrooms ($M = -.07$, $SD = .39$, $p = .006$, 99% CI [.02, .45]. However, there was no significant difference between young men in single-sex and mixed-sex classrooms, $p = .02$, 99% CI [- .42, .01]. Furthermore, when taught in single-sex classrooms, young women ($M = -.30$, $SD = .36$) underperformed relative to their predicted grades compared to young men, who performed in line with their predicted grades ($M = .07$, $SD = .56$), $p < .001$, 99% CI [- .61, - .13]. There was no difference between females

and males in mixed-sex classrooms, $p = .36$, 99% CI [-.12, .26]. When prior attainment was not controlled, there was a main effect of classroom type, $F(1, 259) = .60.77$, $p < .001$, $\eta p^2 = .19$, with students taught in single-sex classrooms ($M = 8.10$, $SD = .93$) seemingly outperforming those taught in mixed-sex classrooms ($M = 6.98$, $SD = 1.13$), $p < .001$, 99% CI [- 1.37, - .68]. Including FSM, EAL and SEN status did not influence these findings.

Discussion

The current study evaluated the efficacy of a single-sex classroom initiative implemented in a co-educational school in the U.K. Such research is able to control for many extraneous environmental variables to a greater extent than research examining the impact of single-sex schooling in different contexts. Overcoming methodological issues within this literature, the current study also controlled for selection effects by accounting for students' previous attainment grades, which were calculated prior to the intervention in line with national curriculum guidelines. In summary, the findings indicate that young women and young men's academic attainment in STEM-related (Mathematics, Science, ICT) and Language subjects (English, foreign Language) did not differ significantly as a function of classroom type. These results are in line with recent meta-analytic findings (Pahlke et al., 2014), which reveal limited evidence for the effectiveness of single-sex classrooms on achievement when controlling for prior achievement.

Findings also indicate that young women underperformed relative to their predicted grades in Non-STEM subjects when they were taught in single-sex relative to mixed-sex classrooms. Additionally, young women taught in single-sex classes underperformed relative to their predicted grades in Non-STEM subjects compared to young men, who performed in line with their predicted grades. This finding contrasts with previous research suggesting that female students may benefit more than males when taught in single-sex compared to mixed-sex classrooms (Alon & Gelbgiser, 2011; Lee & Bryk, 1986; Mullholland et al., 2004). In

order to understand this finding, it may important to reflect on the nature of the school subject or pedagogic context. Specifically, subjects such as art, drama and music are more open-ended by nature and often involve more peer observation and interaction than STEM-related subjects. As a consequence, performance is perhaps more visible in these subjects and females may respond differently to performance appraisal from other ingroup (i.e., a class of other females) relative to outgroup others (i.e., a mixed-class). Furthermore, females may be more self-aware or conscious when participating in performance-based subjects in single-sex groups. In support of this suggestion, research suggests that sex differences in self-concept emerge in adolescence, with girls becoming more self-conscious and aware of criticism than boys (Rankin, Lane, Gibbons, & Gerrard, 2004; Rosenberg & Simmons, 1975), which may help to explain why we found an interaction between gender and the classroom intervention.

When the current results are analyzed without accounting for pre-existing ability, the single-sex classroom initiative appears to be highly efficacious. Such findings are simply a product of the school employing a selective admissions process to assign students to single-sex and mixed-sex classrooms. This research therefore highlights the importance of controlling for selection effects in the evaluation of single-sex classroom initiatives. We argue that studies which do not control for students' prior ability may tell us little about the effectiveness of such interventions.

An additional unexpected finding was that, irrespective of classroom type, male students appeared to underperform relative to their predicted grades in STEM-related subjects compared to female students. Whilst we take caution in inferring explanations from these findings, they may be interpreted in numerous ways. For example, this suggests that males' predicted grades for STEM-related subjects may be overinflated relative to females. This suggestion appears to be supported because young women achieve higher predicted and actual grades compared to young men for all school subjects when analyzing predicted and

obtained grades separately. However, when a difference score is calculated, males' grades in STEM-related subjects appear to be over predicted. Moreover, students are informed typically about their predicted grades in order for teachers to set goals and encourage students to achieve these grades. However, it is plausible that, if predicted grades are set too high, this might have a paradoxical effect on motivation and subsequent exam performance because children feel that their predicted grades are unobtainable. We urge additional research to explore the factors which may explain these pattern of results, and to elucidate whether these findings emerge in other educational settings. Such findings, if corroborated, could have major implications for policy and practice.

Limitations and Future Research Directions

The current research and many previous studies focus on the impact of gender-segregated educational initiatives on academic attainment. As such, there is a lack of research which examines other related psychosocial outcomes that may be influenced by single-sex schooling or classroom interventions. Moreover, studies that do examine additional factors have presented somewhat mixed findings. Although some research indicates that single-sex classrooms may lessen the salience of gender-related stereotypes and performance expectations to bolster students' performance (Elizaga & Markman, 2008; Huguet & Régner, 2007; Inzlicht & Ben-Zeev, 2000; 2003; Picho & Stephens, 2012), other research suggests that gender saliency in single-sex classrooms may exacerbate intergroup biases (Fabes et al., 2015; Halpern et al., 2011; Martin & Fabes, 2001). Accordingly, we recommend that future research examines how single-sex educational strategies may impact on psychological factors such as mindset, competence beliefs, academic self-efficacy, self-esteem, gender stereotyping and intergroup attitudes, in addition to academic attainment. The challenges that may arise

when students subsequently rejoin the opposite sex in post-school settings also warrant further consideration.

Practice Implications

This research proffers both pragmatic and methodological implications. First, we demonstrate how schools assign students typically to educational interventions using selective admissions criteria. Practically, it may be difficult for schools to assign students randomly to single-sex or mixed-sex classrooms because they are taught in ability settings in U.K schools (e.g., grouping students into lower, middle and upper ability groups). In such cases, it is recommended that researchers account for prior achievement in order to elucidate reliably whether single-sex environments represent a practical strategy to bolster academic attainment over co-educational schooling. Studies that do not control for selection effects may tell us little about the effectiveness of such interventions on scholastic achievement.

Second, in this case, the school had implemented a single-sex educational intervention to alleviate a perceived gender-achievement gap in scholastic achievement. Our research allowed us to inform the school whether there were indeed gender differences (separate analyses of obtained grades indicated that females were outperforming males in all school subjects), and whether the single-sex classroom initiative was successful in alleviating these. Given the findings, this evaluation enables the school to examine additional strategies, other than single-sex classroom instruction, that may be more effective in lessening achievement gaps. It also allows them to assess critically whether to continue this single-sex classroom initiative for students entering high school in the future.

Conclusion

This research controlled for students' prior attainment to evaluate the effectiveness of a single-sex classroom initiative implemented in a co-educational, comprehensive U.K school. In summary, findings indicate that young women and young men did not appear to benefit from being taught in single-sex relative to mixed-sex classrooms in Language and STEM-related subjects. Moreover, the single-sex intervention had a seemingly negative impact on young women taught in Non-STEM subjects, who underperformed relative to their predicted grades compared to those taught in mixed-sex classrooms. When prior ability was not controlled for, the intervention appears to be highly efficacious, highlighting the confounding influence of selection effects. These findings therefore demonstrate how the observed advantages of single-sex educational initiatives are reduced greatly when accounting for students' previous scholastic performance. They also provide empirical support for the notion that much of the reported success of gender-segregated education may be attributable to selection effects (Hayes et al., 2011; Signorella, Hayes, & Li, 2013), with this methodological issue distorting the interpretations of research in this area.

References

Alon, S., & Gelbgiser, D. (2011). The female advantage in college academic achievements and horizontal sex segregation. *Social Science Research, 40,* 107-119. doi: 10.1016/j.ssresearch.2010.06.007

Arnot, G., David, M., & Weiner, G. (1999). *Narrowing the gender gap.* Cambridge: Polity Press.

Baker, D. P., Riordan, C., & Schaub, M. (1995). The effects of sex-grouped schooling on achievement: The role of national context. *Comparative Education Review, 39,* 468-482. doi: 10.1086/447341

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education, 15,* 427-448. doi: 10.1007/s11218-012-9185-3

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behavior, 34,* 39-89. doi: 10.1016/S0065-2407(06)80004-2

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's stereotyping and prejudice. *Current Directions in Psychological Science, 16,* 162-166. doi: 10.1111/j.1467-8721.2007.00496.x

Bigler, R., & Signorella, M. L. (2011). Single-sex education: New perspectives and evidence on a continuing controversy. *Sex Roles, 65,* 659-669. doi: 10.1007/s11199-011-0046-x

Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96,* 97-109. doi: 10.1037/0022-0663.96.1.97

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science,* 1-5. doi: 10.1177/0963721410383241

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135,* 218-261. doi: 10.1037/a0014412

Cumming, G., & Calin-Jageman, R. (2017). Introduction to the new statistics: Estimation, open science, & beyond. Routledge: New York.

Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning & Instruction, 39,* 1-10. Doi: 10.1016/j.learninstruc.2015.04.002

Dai, D. Y. (2001). A comparison of gender differences in academic self-concept and motivation between high-ability and average Chinese adolescents. *Journal of Advanced Academics, 13,* 22-32. doi: 10.4219/jsge-2001-361

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science, 314,* 435. doi: 10.1126/science.1131100

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance whilst controlling for test anxiety. *Behavioural and Brain Functions, 8,* 33. doi: 10.1186/1744-9081-8-33

Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. *Journal of Economic Behavior & Organization, 115,* 123-143. doi: 10.1016/j.jebo.2014.08.004

Elizaga, R. A., & Markman, K. D. (2008). Peers and performance: How in-group and out-group comparisons moderate stereotype threat effects. *Current Psychology, 27,* 290-300. doi: 10.1007/s12144-008-9041-y

Else-Quest, N. M., & Peterca, O. (2015). Academic attitudes and achievement in students of urban public single-sex and mixed-sex high schools. *American Educational Research Journal, 52,* 693-718. doi: 10.3102/0002831215591660

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103-127. doi: 10.1037/a0018053

Fabes, R. A., Lynn-Martin, C., Hanish, L. D., Galligan, K., & Pahlke, E. (2015). Gender-segregated schooling: A problem disguised as a solution. *Educational Policy, 29,* 431-447. doi: 10.1177/0895904813492382

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53,* 25-44. doi: 10.1016.jsp.2014.10.002

Gillibrand, E., Robinson, P., Brawn, R., & Osborn, A. (1999). Girls' participation in physics in single sex classes in mixed schools in relation to confidence and achievement. *International Journal of Science Education, 21,* 349-362. doi: 10.1080/095006999290589

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 1,* 1-51. doi: 10.111/j.1529-1006.2007.00032.x

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., . . . & Lynn Martin, C. (2011). The pseudoscience of single-sex schooling. *Science, 23,* 1706-1707. doi: 10.1126/science.1205031

Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development, 84,* 1716-1733. doi: 10.1111/cdev.12079

Hayes, A. R., Pahlke, E., & Bigler, R. (2011). The efficacy of single-sex education: Testing for selection and peer quality effects. *Sex Roles, 65,* 693-703. doi: 10.1007/s11199-010-9903-2

Hoffman, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning & Instruction, 12,* 447-465. doi: 10.1016/S0959-4752(01)00010-X

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99,* 545-560. doi: 10.1037/0022-0663.99.3.545

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11,* 365-371. doi: 10.1111/1467-9280.00272

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95,* 796-805. doi: 10.1037/0022-0663.95.4.796

Jackson, C., & Smith, I. D. (2000). Poles apart? An exploration of single-sex and mixed-sex educational environments in Australia and England. *Educational Studies, 26,* 409-42. doi: 10.1080/03055690020003610

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105,* 198-214. doi: 10.1037/0033-2909.105.2.198

Lee, V. E., & Bryk, A. S. (1986). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 78,* 381-395. doi: 10.1037/0022-0663.78.5.381

Lee, V. E., & Marks, H. M. (1990). Sustained effects of the single-sex secondary school experience on attitudes, behaviors, and values in college. *Journal of Educational Psychology, 82,* 578-592. doi: 10.1037/0022-0663.82.3.578

Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). *Single-sex versus coeducational schooling: A systematic review.* (No. 2005-01). Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Retrieved from https://www2.ed.gov/rschstat/eval/other/single-sex/single-sex.pdf

Marsh, H. (1989). The effects of attending single-sex and Catholic coeducational high schools on achievement, attitudes and behaviours and on sex differences. *Journal of Educational Psychology, 81,* 70-85. doi: 10.1037/0022-0663.81.1.70

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

Miller, D. I., Eagley, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107,* 631-644. doi: 10.1037/edu0000005

Mullholland, J., Hansen, P., & Kaminski, E. (2004). Do single-gender classrooms in coeducational settings address boys' underachievement? An Australian study. *Educational Studies, 30,* 19-32. doi: 10.1080/0305569032000159714

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic perspectives, 24,* 129-144. doi: 10.1257/jep.24.2.129

Pahlke, E., & Hyde, J. S. (2016). The debate over single-sex schooling. *Child Development Perspectives, 10,* 81-86. doi: 10.1111/cdep.12167

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin, 140,* 1042-1072. doi: 10.1037/a0035740

Pahlke, E., Hyde, J. S., & Mertz, J. E. (2013). The effects of single-sex compared with coeducational schooling on mathematics and science achievement: Data from Korea. *Journal of Educational Psychology, 105,* 444-452. doi: 10.1037/a0031857

Pajares, F., & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology, 26,* 366–381. doi: 10.1006/ceps.2000.1069

Park, H., Behrman, J. R., & Choi, J. (2013). Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in Seoul high schools. *Demography, 50,* 447-469. doi: 10.1007/s13524-012-0157-1

Parker, L. H., & Rennie, L. J. (2002). Teachers' implementation of gender-inclusive instructional strategies in single-sex and mixed-sex classrooms. *International Journal of Science Education, 24,* 881-897. doi: 10.1080/09500690110078860

Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology, 65,* 26-30. doi: 10.1016/j.jesp.2016.02.008

Picho, K., & Stephens, J. M. (2012). Culture, context and stereotype threat: A comparative analysis of young Ugandan women in coed and single-sex schools. *The Journal of Educational Research, 105,* 52-63. doi: 10.1080/00220671.2010.517576

Rankin, J. L., Lane, D. J., Gibbons, F. X., & Gerrard, M. (2004). Adolescent self-consciousness: Longitudinal age changes and gender differences in two cohorts. *Journal of Research on Adolescence, 14,* 1-21.

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107*, 645-662. doi: 10.1037/edu0000012

Rosenthal, L., London, B., Levy, S. R., & Lobel, M. (2011). The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university. *Sex Roles, 65,* 725-738. doi: 10.1007/s11199-011-9945-0

Rosenberg, F. L., & Simmons, R. G. (1975). Sex differences in the self-concept in adolescence. *Sex Roles, 1*, 147-159.

Schneider, F. W., & Coutts, L. M. (1982). The high school environment: A comparison of coeducational and single-sex schools. *Journal of Educational Psychology, 74,* 898-906. doi: 10.1037/0022-0663.74.6.898

Shapka, J. D. (2009). Trajectories of math achievement and perceived math competence over high school and postsecondary education: Effects on an all-girl curriculum in high school. *Educational Research and Evaluation: An International Journal on Theory and Practice, 15,* 527-541. doi: 10.1080/13803610903354775

Shapka, J. D., & Keating, D. P. (2003). Effects of a girls-only curriculum during adolescence: Performance, persistence, and engagement in mathematics and science. *Educational Research Journal, 40,* 929-960. doi: 10.3102/00028312040004929

Signorella, M. L., & Bigler, R. S. (2013). Single-sex schooling: Bridging science and school boards in educational policy. *Sex Roles, 69*(7-8), 349-355. doi: 10.1007/s11199-013-0313-0

Signorella, M. L., Hayes, A. R., & Li, Y. (2013). A meta-analytic critique of Mael et al.'s (2005) review of single sex schooling. *Sex Roles, 69,* 423-441. doi: 10.1007/s11199-013-0288-x

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28. doi: 10.1006/jesp.1998.1373

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613-629. doi: 10.1037/0003-066X.52.6.613

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS One,* e57988. doi: 10.1371/journal.pone.0057988

Sullivan, A. (2009). Academic self-concept, gender, and single-sex schooling. *British Educational Research Journal, 35,* 259-288. doi: 10.1080/01411920802042960

Sullivan, A., Joshi, H., & Leonard, D. (2010). Single-Sex schooling and academic attainment at school and through the lifecourse. *American Educational Research Journal, 47,* 6-36. doi: 10.3102/0002831209350106

The National Curriculum. (2010). *The National Curriculum: Level descriptions for subjects*. Coventry: The Qualifications and Curriculum Authority (QCA). Available from: http://www.empiribox.org/wp-content/uploads/2014/08/Level-Descriptors-for-Science.pdf

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics, 50,* 49-62. doi: 10.1023/A:1020518104346

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174-1204. doi: 10.1037/a0036620

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38,* 412-423. doi: 10.1016/j.intell.2010.04.006

Warrington, M., & Younger, M. (2001). Single-sex classes and equal opportunities for girls and boys: Perspectives through time from a mixed comprehensive school in England. *Oxford Review of Education, 27,* 339-356. doi: 10.1080/03054980120067393

Warrington, M., & Younger, M. (2003). 'We decided to give it a twirl': Single-sex teaching in English comprehensive schools. *Gender and Education, 15,* 339-350. doi: 10.1080/09540250310001610553

Wood, D., Kurtz-Costes, B., Rowley, S. J., & Okeke-Adeyanju, N. (2010). Mothers' academic gender stereotypes and education-related beliefs about sons and daughters in African American families. *Journal of Educational Psychology, 102,* 521-530. doi: 10.1037/a0018481

Younger, M., & Warrington, M. (2002). Single-sex teaching in a co-educational comprehensive school in England: An evaluation based upon students' performance and classroom interactions. *British Educational Research Journal, 28,* 353-374. doi: 10.1080/01411920220137449

Younger, M. R. & Warrington, M. (2006). Would Harry and Hermione have done better in single-sex classes? A review of single-sex teaching in coeducational secondary schools in the United Kingdom. *American Educational Research Journal, 43,* 579-620. doi: 10.1302/00028312043004579

Young-Suk, K., Al Otaiba, S., Wanzek, J., & Brandy, G. (2015). Toward and understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology, 107,* 79-95. doi: 10.1037/a0037210

Table 1

Subject-specific attainment levels based on National Curriculum guidelines, re-coded into ordinal classifications.

	Classification/Grouping		
Levels	Lower	Middle	Upper
Level 2	1	2	3
Level 3	4	5	6
Level 4	7	8	9
Level 5	10	11	12
Level 6	13	14	15
Level 7	16	17	18
Level 8	19	20	21

Table 2

Descriptive statistics for student's academic attainment (controlling for prior achievement) by gender and classroom type within subject areas.

	Subject Areas								
	(a) Languages			(b) STEM			(c) Non-STEM		
	Classroom Type		Gender	Classroom Type		Gender	Classroom Type		Gender
Students' Gender	Single-sex *M* (*SD*)	Mixed-sex *M* (*SD*)	Main Effect *M* (*SD*)	Single-sex *M* (*SD*)	Mixed-sex *M* (*SD*)	Main Effect *M* (*SD*)	Single-sex *M* (*SD*)	Mixed-sex *M* (*SD*)	Main Effect *M* (*SD*)
Young Women	- .36 (.75)	- .10 (.65)	- .21 (.71)	- .32 (.95)	- .24 (.83)	- .27 (.88)[a]	- .30 (.36)[bc]	- .07 (.39)[b]	- .17 (.39)
Young Men	- .13 (.22)	- .21 (.68)	- .18 (.57)	- .61 (.72)	- .57 (.96)	-.58 (.89)[a]	.07 (.56)[c]	- .14 (.50)	- .07 (.53)
Classroom Main Effect	- .26 (.59)	- .16 (.67)		- .45 (.92)	- .44 (.92)		- .14 (.49)	- .11 (.46)	

Note. Languages includes English and foreign languages; STEM includes Science, Mathematics and Information and Communication Technology; Non-STEM includes art, drama and music. Different subscripts comparing means for the main effects of classroom type and gender, as well as for the Classroom type*Gender interaction, indicate a statistically significant difference ($p < .01$).

Supporting Information File 1

The decision was made to compute a composite grade for STEM-subjects (mathematics, science, ICT), Non-STEM subjects (art, drama, music) and Languages (English, foreign language) within the reported study findings to control for Type 1 error rates. Separate one-way Analysis of Variance analyses for each subject type, with an adjusted *p*-value of .01, are as follows:

STEM Subjects

Science. There was no significant main effect of classroom type on students' attainment in science, $F(1, 175) = .31$, $p = .58$, $\eta_p^2 = .002$.

Mathematics. There was no significant main effect of classroom type on students' attainment in mathematics, $F(1, 256) = 1.20$, $p = .27$, $\eta_p^2 = .005$.

Information and Communications Technology (ICT). There was no significant main effect of classroom type on students' attainment in ICT, $F(1, 258) = .06$, $p = .81$, $\eta_p^2 < .001$.

Art. There was no significant main effect of classroom type on students' attainment in art, $F(1, 256) = 2.95$, $p = .09$, $\eta_p^2 = .01$.

Drama. There was a significant main effect of classroom type on students' attainment in drama, $F(1, 259) = 7.50$, $p = .007$, $\eta_p^2 = .03$. Students taught in single-sex classrooms ($M = -.38$, $SD = .91$) unperformed relative to their predicted grades compared to those taught in mixed-sex classrooms ($M = -.09$, $SD = .78$), p = .007, 99% CI [.02, .57].

Music. There was no significant main effect of classroom type on students' attainment in music, $F(1, 256) = .02$, $p = .89$, $\eta_p^2 < .001$.

English. There was no significant main effect of classroom type on students' attainment in English, $F(1, 258) = .20$, $p = .66$, $\eta_p^2 = .001$.

Foreign language. There was no significant main effect of classroom type on students' attainment in foreign language, $F(1, 251) = 1.80$, $p = .18$, $\eta_p^2 = .007$. See Table 1 for summary of descriptive statistics.

Supporting Table 1.

Descriptive statistics for student's academic achievement across separate school subjects (controlling for prior achievement) as a function of classroom type.

	Classroom Type	
	Single-sex	Mixed-sex
School Subject		
Science	- .29 (1.66)	- .46 (1.84)
Mathematics	- .16 (.76)	- .29 (.91)
ICT	- .65 (1.27)	- .61 (1.37)
Art	- .01 (1.11)	- .21 (.80)
Drama	- .38 (.91)[a]	- .09 (.78)[a]
Music	- .02 (.32)	- .01 (.52)
English	- .16 (.47)	- .12 (.94)
Foreign language	- .34 (.98)	- .19 (.81)

Note. Different subscripts indicate a statistically significant difference, $p < .01$.