

Journal of Educational Psychology

Gender Cognitions Before and After Graduating From Single-Sex Versus Coeducational High Schools: A Longitudinal Study Using Propensity Score Matching

Sylvia Yun Shi, Gu Li, Lynn S. Liben, Zhansheng Chen, and Wang Ivy Wong

Online First Publication, July 18, 2024. <https://dx.doi.org/10.1037/edu0000879>

CITATION

Shi, S. Y., Li, G., Liben, L. S., Chen, Z., & Wong, W. I. (2024). Gender cognitions before and after graduating from single-sex versus coeducational high schools: A longitudinal study using propensity score matching.. *Journal of Educational Psychology*. Advance online publication. <https://dx.doi.org/10.1037/edu0000879>

Gender Cognitions Before and After Graduating From Single-Sex Versus Coeducational High Schools: A Longitudinal Study Using Propensity Score Matching

Sylvia Yun Shi^{1, 2}, Gu Li³, Lynn S. Liben⁴, Zhansheng Chen⁵, and Wang Ivy Wong^{1, 2}

¹Gender Studies Programme, The Chinese University of Hong Kong

²Department of Psychology, The Chinese University of Hong Kong

³Faculty of Arts and Sciences, NYU Shanghai

⁴Department of Psychology, The Pennsylvania State University

⁵Department of Psychology, The University of Hong Kong

Comparisons between single-sex (SS) and coeducational (CE) schools are often limited in design. Associations between SS schooling and gender cognitions remain especially unclear. We compared gender cognitions in students from SS versus CE high schools. A longitudinal design addressed long-term reciprocal effects, and propensity score matching addressed preexisting (confounding) differences between SS and CE students. During the final year of high school ($N = 667$) and again following graduation ($N = 463$), students completed measures of gender salience, gender stereotypes, felt pressure for gender conformity, and exposure to gender equality. Before graduation, SS students were more gender salient about self, reported more exposure to gender equality, and had less pressure for gender conformity. The magnitudes of these SS–CE differences were small but did not change significantly over time. However, gender salience about self did not predict or mediate other gender cognitions as often hypothesized, and there were no differences in gender salience about others or in gender stereotypes. We conclude that SS schooling is related to some gender-related cognitions both prior to and following high school graduation. However, findings are neither consistently supportive nor critical of either type of schooling. SS schooling was positively related to students' attention to their own gender, but how this attention relates to other gender cognitions may vary across individuals in ways that need further research. Our design innovations and inclusion of debated yet understudied outcomes contribute to a fuller evaluation of SS versus CE schooling by examining long-term development and by expanding the outcomes evaluated.


Educational Impact and Implications Statement

We employed a novel rigorous design that followed students longitudinally before and after their high school graduation and examined understudied outcomes (i.e., gender cognitions) to evaluate single-sex versus coeducational schooling more fully. Single-sex, compared to coeducational, high school students are slightly more conscious of their gender but are no more or are less gender-typed in some other gender cognitions. High school students' gender cognitions undergo temporal changes regardless of school type.

Keywords: single-sex schooling, gender salience, gender stereotypes, gender equality, cross-lagged design

Supplemental materials: <https://doi.org/10.1037/edu0000879.supp>

Beth Kurtz-Costes served as action editor.

Wang Ivy Wong  <https://orcid.org/0000-0002-9428-2832>

This study was supported by a General Research Fund awarded to Wang Ivy Wong by the Hong Kong Research Grants Council (17610818). The data, codes, and measures of the current study are available from the corresponding author upon reasonable request.

Sylvia Yun Shi served as lead for formal analysis, validation, and visualization, contributed equally to investigation and project administration, and served in a supporting role for methodology, writing–original draft, and writing–review and editing. Gu Li contributed equally to formal analysis and served in a supporting role for funding acquisition, investigation, methodology, supervision, validation, writing–original draft, and writing–review and editing.

Lynn S. Liben served in a supporting role for conceptualization, funding acquisition, methodology, and writing–review and editing. Zhansheng Chen served in a supporting role for funding acquisition, methodology, and writing–review and editing. Wang Ivy Wong served as lead for conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, writing–original draft, and writing–review and editing and served in a supporting role for formal analysis and validation. Sylvia Yun Shi and Wang Ivy Wong contributed equally to data curation.

Correspondence concerning this article should be addressed to Wang Ivy Wong, Gender Studies Programme, The Chinese University of Hong Kong, 248, Sino Building, Shatin, New Territories, Hong Kong. Email: iwwong@cuhk.edu.hk

Gender segregation prevails across the life span as preferences for own-gender groups begin as early as the preschool years and continue across life. It reinforces gender differences in conversational styles and activity preferences, strengthens ingroup–outgroup biases, supports gender-based power asymmetries in dating scripts, and promotes further segregation in interpersonal relationships (Leaper, 2022; Mehta & Strough, 2009). A common institutional form of gender segregation occurs in schools in which boys and girls are segregated into single-sex (SS) schools or classes. This form of institutional gender segregation exists at all levels of education globally and may have far-reaching developmental impacts (Leaper, 2022; Liben, 2015; Pahlke & Hyde, 2016). Although the popularity of SS education has declined over time in some countries, it remains common in others. For example, 24%–31% of independent schools in Britain and Australia are SS (Cooper, 2010). In the United States, a legal revision in 2006 led the number of SS public schools to surge (Liben, 2015) from two in 1996 to 100 in 2014 (Riordan, 2011). Some Chinese cities have also resorted to all-boys programs in an attempt to improve boys' performance and manhood (Agence France-Presse, 2013). This revival, coupled with uncertainty over the consequences of SS education, has resulted in highly visible debates among SS education practitioners and researchers (Halpern et al., 2011; Liben, 2015; Pahlke & Hyde, 2016; Sax, 2011; Signorella & Bigler, 2013), and in legal battles that saw the closure of many SS classes or schools in U.S. public education (Brown, 2013). Nevertheless, SS schools continue to exist across the globe and gender segregation is common within otherwise coeducational (CE) schools (e.g., in physical or health education classes). Thus, it is unsurprising that comparisons of SS versus CE education remain of great interest to academicians and the public (Robinson et al., 2021).

The current study fills in important research gaps by empirically comparing SS and CE school students on gender cognitions. We collected longitudinal data to allow us to compare SS and CE systems at two points in students' educational careers. The first wave of data collection was scheduled during students' final year of high school and the second was roughly 1.5 years following graduation. This schedule thus bridged an important social transition that differed for SS versus CE students with respect to whether there was, or was not, a dramatic change in the gendered nature of their daily interactions. Thus, our design addressed (a) group comparisons between SS and CE students, (b) within-person comparisons across time before and after graduation and whether these comparisons were moderated by SS–CE group membership, and (c) whether individual's varying levels of gender salience—that is, the degree to which gender is readily accessible, made conscious, and relevant to an individual's cognitive schemas or when making classifications—was related to SS–CE differences in other cognition outcomes.

Apart from providing valuable empirical data relevant for a theoretical debate about links between school structure (SS vs. CE) and gender cognitions such as gender salience, findings that increase understanding of these three issues will contribute to educators' increased understanding of the consequences of differing school structures and ultimately help educators design educational interventions that take gendered structure into account.

It is important to note here that the term “gender salience” has been used in various ways by previous researchers, and no one meaning or measure of the label or concept is universally endorsed. Some research has focused on gender salience as a person variable.

Illustrative is the distinction between gender schematic versus gender aschematic people made by Bem (1981) and the distinction between children with stronger versus weaker gender salience filters proposed by Liben and Bigler (2002). Other work has focused on gender salience as an environmental variable, for example, distinguishing between environments that highlight versus ignore gender in language (Bigler & Leaper, 2015; Palomares, 2008), in classroom gender composition (Fabes et al., 2015, 2018), or in laboratory experiments that manipulated ad hoc group gender composition (Cota & Dion, 1986). Yet other work has emphasized the inseparable or “relational” importance of person-centered and context-centered constructs, an approach exemplified in developmental intergroup theory (DIT; Bigler & Liben, 2006, 2007; see Liben, 2014). Consistent with DIT, in various works developed from the distinctiveness theory (McGuire, 1984), gender salience is conceptualized as showing contextual differences along factors such as group proportion and distinctive traits relative to other people in a given situation (environmental variable; e.g., Cota & Dion, 1986; McGuire & Padawer-Singer, 1976; McGuire et al., 1979). However, the situation can be both transient (e.g., ad hoc group gender composition; Cota & Dion, 1986) or chronic or relatively long-term (e.g., family gender composition and classroom gender composition; McGuire & Padawer-Singer, 1976; McGuire et al., 1979). Chronic situations can give rise to individual differences in gender salience (e.g., McGuire et al., 1979; Wood & Eagly, 2015).

It would be beyond the scope of a single empirical article to review the many past uses of the term gender salience. It would be valuable if future work clarified and compared interpretations of gender salience in more detail. For the current article, we draw our use of gender salience from related research stemming from social identity theory and cognitive developmental theory (including the DIT) and distinctiveness theory that define gender salience as the extent to which someone uses gender for making membership categorizations or judgments, reflecting their consciousness of and attention to gender (Bigler & Liben, 2006, 2007; Cota & Dion, 1986; Liben, 2017; McGuire, 1984; McGuire & Padawer-Singer, 1976; McGuire et al., 1979; Mehta, 2015; Palomares, 2008; Serbin & Sprafkin, 1986; Wong et al., 2018). Because situational factors such as the gender composition of one's school are relatively chronic and because person-centered and context-centered constructs are inseparable (Bigler & Liben, 2006; see Liben, 2014), we conceptualize gender salience as both a person attribute and a situational variable (see also Wood & Eagly, 2015).

The Debate About Gender Salience in SS Versus CE Schools

Recent reviews have questioned the alleged academic advantages of SS over CE education, arguing that advantages previously reported for SS schools instead resulted largely from differences in student socioeconomic status and prior levels of academic achievement (Pahlke, Hyde, & Allison, 2014; Signorella et al., 2013). Those who criticize SS education often assert that SS settings will lead students to have higher gender salience, that is, to be more attentive to gender than their CE counterparts. However, the assertion that gender salience is greater among SS students has rarely been tested empirically in actual school settings. It has been theorized that highly gender-salient people view others through gendered lenses and make more gender-based categorizations (Bigler & Liben, 2006, 2007; Liben & Bigler, 2002, 2017), possibly intensifying gender segregation and reinforcing gender stereotypes. There have been experimental studies designed to

model highly gender-salient settings. For example, when teachers create gender-salient environments (e.g., use more gender-specific language or asking boys and girls to line up separately), children develop stronger gender stereotypes, interact less with other-gender peers, and view other-gender peers less positively (Hilliard & Liben, 2010). Gender salience may also have broader impacts, including deterring individuals from pursuing courses and careers that are perceived as inappropriate for their own gender (Bigler & Liben, 2006, 2007). A contradictory hypothesis is given by advocates of SS schooling (e.g., Sax, 2011). For example, advocates posit that SS environments reduce gender salience because there is no opportunity for teachers to treat boys and girls differentially within the same school or classroom as all students are of the same gender. Proponents of SS schools also argue that these contexts provide increased exposure to gender equality, especially in all-girls schools. In contrast, critics of SS contexts (e.g., Halpern et al., 2011; Liben, 2015; Pahlke & Hyde, 2016) posit that SS schools will increase gender salience because segregation constantly reminds students that gender is important.

Several social constructivist theories are relevant for understanding how SS and CE students might differ in gender cognitions. First, DIT, pioneered by Bigler and Liben (2006, 2007; see also Liben, 2017), posits that an identity (gender in our case) becomes salient (cognitively accessible and relevant) under four environmental conditions—when groups are perceptually discriminable, in the minority, explicitly labeled, and implicitly used—and furthermore, that a higher degree of salience leads to greater bias. Many researchers use DIT to hypothesize that, owing to some characteristics of SS schools (e.g., gender labels in school names), SS students are more gender-salient and stereotyped (Halpern et al., 2011; Liben, 2015; Pahlke, Hyde, & Allison, 2014). This hypothesis was contradicted by a meta-analysis finding reduced gender-stereotyping in SS girls; however, the authors noted that heterogeneity of measures made this finding hard to interpret and there was a lack of controlled studies of SS boys (Pahlke, Hyde, & Allison, 2014). Others (e.g., Sax, 2011), emphasizing other characteristics of SS schools (e.g., less differential treatment, own gender being the majority at school), predict the opposite, that is, lower gender salience in SS students. Both critics and supporters of SS schooling cite research on how people perform in same- versus mixed-gender settings as supporting evidence (e.g., Halpern et al., 2011; Sax, 2011), but such research does not directly assess gender salience in SS and CE students. Therefore, even when researchers and stakeholders agree with the principles of DIT, they may still disagree about the relative gender salience of SS and CE students because SS schools have features that both increase and decrease gender salience.

Second, social-cognitive theories, which emphasize differential treatment and gender schemas (Bussey & Bandura, 1999; Cook et al., 2022), are relevant to hypothesizing outcomes on intergroup stereotypes. In peer socialization studies, same-gender peers reinforce gender-typed behaviors and attitudes (Mehta & Strough, 2009). Hence, SS students may be more gender-stereotyped because of reduced socialization by other-gender peers, and they may become less so after graduation because of increased socialization of this kind. This prediction is consistent with intergroup contact theory, which suggests that intergroup contact reduces outgroup prejudice (Pettigrew & Tropp, 2006). However, social-cognitive theories can also make the opposite prediction insofar as the presence of both genders increases opportunities for others to apply differential socialization and for people to make gender-based social comparisons (McHale et al., 2003).

Existing Findings on Gender Cognitions in SS Versus CE Students

Few studies examined gender cognitions in SS and CE students and even fewer within these addressed gender salience. Available findings are mixed. For example, one study found SS girls to feel more gender-typical and pressured to conform to gender norms than CE girls (Drury et al., 2013), but another found girls randomly assigned to SS Physics classes to have lower endorsement of, and longer time responding to, feminine traits. The latter finding was interpreted as showing SS girls have less accessible gender-related self-knowledge and, thus, lower gender salience (Kessels & Hannover, 2008). However, the outcome measure used is not well matched to gender salience as we have defined it here which is readiness and propensity for using gender in making categorizations (Cota & Dion, 1986; McGuire, 1984; McGuire & Padawer-Singer, 1976; McGuire et al., 1979; Palomares, 2008; Wong et al., 2018). Lower endorsement of feminine traits by the girls in SS classes may reflect an outcome of gender salience instead of gender salience per se. A prior cross-sectional study (Wong et al., 2018) used a traditional measure of gender salience and found that SS high school students were more likely than CE high school students to mention gender when asked who they are (or are not) even when demographic covariates were considered. This finding provides initial direct evidence for the hypothesis that gender is more salient in SS schools (Halpern et al., 2011; Liben, 2015; Pahlke, Hyde, & Allison, 2014). Nevertheless, this finding requires replication in a well-controlled longitudinal study.

Design Challenges in Prior Studies

Given the mixed predictions and limitations in studies on whether and how SS and CE students differ on gender cognitions, more empirical evidence is needed. It would be informative and timely for studies to focus on understudied gender cognition variables, especially gender salience, and to assess these outcomes longitudinally and to control for background differences between SS and CE students. However, ethical and practical restrictions mean that it is rarely possible to assign students randomly to SS versus CE schools. As a result, attributes of the students who are ultimately enrolled in the two kinds of schools can be expected to differ. For example, students' from SS schools tend to have higher socioeconomic backgrounds and academic motivations (Bigler et al., 2014; Halpern et al., 2011; Nagengast et al., 2013; Pahlke, Hyde, & Allison, 2014; Signorella & Bigler, 2013; Signorella et al., 2013).

Thus, across all investigations of SS schooling (whether focused on gender cognitions or otherwise), several methodological constraints and considerations apply. Truly randomized experiments that allocate students randomly to SS and CE schools for research purposes are extremely rare. In nonexperimental research, many researchers failed to include comparison groups and made no attempt to address potential self-selection bias (i.e., students may already differ upon school entry; Pahlke, Hyde, & Allison, 2014).

Studies that statistically control for differences between SS and CE students (e.g., socioeconomic status, religion) or control for prior performance (e.g., by assessing academic performance both before and after a year of SS vs. CE instruction) are more effective in examining the impact of SS versus CE schooling (Liben, 2015; Pahlke, Hyde, & Allison, 2014). For students' academic outcomes, meta-analyses (Pahlke, Hyde, & Allison, 2014; Signorella et al., 2013) have

shown that the inclusion of even a few demographic covariates is sufficient to eliminate inaccurate conclusions (false positives) about significant advantages or disadvantages of a given school type. Studies adopting additional strategies to make the controlling even more stringent are rare. A few earlier investigators have used propensity score matching to statistically match SS and CE students to create data sets resembling an experimental design (e.g., Garcia-Gracia & Donoso Vázquez, 2016; Gee & Cho, 2014; Nagengast et al., 2013; Paredes, 2022) and some quasi-experimental studies in South Korea were able to capitalize on an exceptional policy that led to nearly random assignments of students to SS versus CE schools within school districts (e.g., Lee & Nakazawa, 2022; Lee & Sambanis, 2023).

Despite increased attention to problems of nonrandom assignment in recent studies, well-controlled studies are still rare. Furthermore, the majority of SS versus CE studies still focus on academic and career achievement or academic self-concepts (Mael et al., 2005; Pahlke, Hyde, & Allison, 2014; Robinson et al., 2021). Few focused on gender cognitions. A rare exception was a study by Kessels and Hannover (2008), who studied accessibility of gender self-knowledge using random assignment. However, assignment to SS versus CE occurred at the class level. In addition, they assessed gender salience by asking students to indicate the degree to which they believed that various gender-typical descriptions applied to themselves rather than by assessing their awareness of gender when making categorizations. Although Lee and Sambanis (2023) investigated some aspects of gender cognitions (i.e., hostile sexism and gender role attitude) in adult SS and CE graduates using randomized sampling (and found no difference), other critical gender cognition outcomes such as gender salience, felt pressure to conform to gender stereotypes, or exposure to gender equality require more well-controlled study.

Also, very few studies addressed the longitudinal changes of SS and CE students. Some studies utilizing large data sets are limited in the types and depth of outcome measures used posthigh school graduation. For example, many past studies commonly tapped into status variables such as educational or career attainment and choices (Lee & Nakazawa, 2022; Nagengast et al., 2013; Sullivan et al., 2010). The few studies to date that followed up on the nonachievement-related outcomes of SS and CE students have focused on marriage (e.g., Sullivan et al., 2012). One study of girls in a school with a CE campus and an SS campus transitioning to CE found reduced stereotyping over a year regardless of whether students were in CE classes or SS classes (Signorella et al., 1996). Gender stereotypes in adulthood have been shown to be similar in SS and CE graduates who had been randomly assigned to SS or CE high schools (Lee & Sambanis, 2023), but like the other studies, measurement was only made at one time. Therefore, more studies are needed to reveal how SS students adapt at transition windows.

In sum, educators and researchers have been debating the consequences of SS versus CE schooling, but the existing literature has several important limitations. First, irrespective of the outcome being studied, most comparisons of students in SS versus CE schools have been confounded by differences in student background. Second, most studies have focused on a limited range of student outcomes, most assessing academic achievement. Third, very few investigations have measured gender salience directly despite the fact that this construct has often been viewed as at the crux of the SS versus CE debate. Wong et al. (2018) compared the gender salience of SS and CE students but even that study was limited insofar as it was cross-sectional rather than longitudinal. As explained next, our work was designed to

build from and extend past work. We did so by examining understudied gender cognition outcomes including gender salience and the longitudinal changes of SS and CE students before and after they left the gender-segregated environment and entered a mixed-gender world. Propensity score matching and repeated measures were administered to address some of the critical methodological limitations discussed above (i.e., confounding background differences between SS and CE students and lack of longitudinal studies).

Design Innovations: Propensity Score Matching and Cross-Lagged Design

Propensity score matching can model the assignment of students into different school types. Although matching does not permit conclusions about causality like those from randomized control trials, it does permit stronger conclusions about observed associations than do designs that ignore confounding group differences (Nagengast et al., 2013). Specifically, propensity score matching can closely mimic a randomized experiment (Stuart, 2010). A good balance between sample characteristics can be achieved before analyzing the outcomes (Rubin, 2005). Propensity score matching performs better than regression methods such as analysis of covariance especially when there is insufficient overlap between the treatment and the control groups in the covariate distributions (Stuart, 2010). Although earlier researchers have recommended propensity score matching for studying the effects of SS schooling (Nagengast et al., 2013; Pahlke, Hyde, & Allison, 2014), very few investigators have applied it, and those who did (e.g., Garcia-Gracia & Donoso Vázquez, 2016; Gee & Cho, 2014; Nagengast et al., 2013; Paredes, 2022) did not focus on gender cognition outcomes.

In addition, a cross-lagged longitudinal design can examine the differences between SS and CE schooling over time around high school graduation and the potential mediation effects of gender salience. Unlike the basic regression approaches or group comparison methods more commonly used in the SS schooling literature, cross-lagged panel (CLP) model enables testing for prospective associations that may be reciprocal. Also, this longitudinal approach is helpful for probing the potential factors that mediate the predictive associations between school type and students' long-term outcomes so that initial insights can be gained into the underlying mechanisms.

The Current Study

In the current study, we empirically compared SS and CE students before and after their high school graduation on gender salience, gender stereotypes, felt pressure for gender conformity, and exposure to gender equality, and we examined associations among these variables. Combining cross-lagged longitudinal design and propensity score matching is a novel approach for studying SS schooling and provides a more stringent test of the differences between SS and CE schooling outcomes.

The period between the final year of high school and a year post-graduation was chosen for two reasons. First, almost all local kindergartens are CE, and SS schools are very few relative to CE schools at the primary (about 6% SS) and secondary school (about 16% SS) levels, making it very difficult to conduct a longitudinal study that could follow up a large sample of CE students entering SS schools at these lower levels of education. On the other hand, tertiary schools are all CE in Hong Kong. Thus, the high school graduation period is

the only transition period in which almost all SS school students recruited initially could be expected to enter a CE environment, making it feasible to recruit a large and balanced sample split by SS and CE students and to examine the effect of change in gender segregation environment. Second, SS and CE students would be expected to show more apparent school-type differences in the final year than earlier years because they have spent more extended time in different schooling environments.

As implied above, our sample came from Hong Kong, which is a former British colony in which the education system and gender differences are similar to those found in the West (e.g., Wong & VanderLaan, 2020). As in the West, SS schools in Hong Kong are also of higher status, are more academically selective, and are often established by churches and missionaries originating from North America and Western Europe since the 19th century. The initial mission for boys' schools to preach was somewhat unsuccessful and overtaken by the Chinese pragmatic concerns for boys and men to receive academic training in order to obtain respectable jobs (W. Y. Ho, 2004). For girls' schools, many were set up with the liberal aim of helping disadvantaged girls become independent women. Following the modern ideology of equal opportunities, there has been a deemphasis of gender orientation in education (W. Y. Ho, 2004). However, due to their higher prestige and more international background, SS schools are still coveted in Hong Kong. Hong Kong culture emphasizes academic achievement; local high schools are assigned a band to indicate their academic standing (Band 1 being the highest and Band 3 being the lowest). The education culture is highly merit-based and discipline-focused, although there are also classes on culture, sports, and art, and most students take up extracurricular activities for further learning or as hobbies.

High school allocation is determined by a host of factors, including merit factors such as academic performance and extracurricular participation, as well as nonmerit factors such as district of residence, self-selected preference, religious background, schools' discretion, and partial random allocation. Both gender egalitarian and traditional gender conservative ethos coexist in SS schools as well as CE schools. For example, there is evidence of some feminist voice in SS girls' schools and girls' schools often emphasize training independent women and cultivating middle-class ladies (W. Y. Ho, 2004). Across all schools, certain subjects (e.g., home economics) are subtly offered and designed on the basis of gender differences despite the explicit syllabus and objectives are often portrayed in a gender-neutral manner (W. Y. Ho, 2004).

Based on theories and empirical findings drawn from the West, which we reason to be largely generalizable to Hong Kong, we tested three sets of hypotheses on gender cognitions, which are focused first on differences between school types (Hypothesis 1 [H1]), second on longitudinal changes surrounding high school graduation (Hypothesis 2 [H2]), and third on cross-lagged effects among the study variables (Hypothesis 3 [H3]).

H1: School-Type Differences

Although past findings on SS–CE differences in gender stereotypes are mixed and few studies have reported on felt pressure, a cross-sectional study by Wong et al. (2018) has directly examined gender salience in SS and CE students. Thus, we formulated our hypotheses on school-type differences based on this earlier study (Wong et al., 2018) and on processes identified in DIT (Bigler & Liben, 2006,

2007) and distinctiveness theory (Cota & Dion, 1986; McGuire, 1984; McGuire et al., 1979; McGuire & Padawer-Singer, 1976). We hypothesized that prior to high school graduation, SS students would be more gender salient (Hypothesis 1a [H1a]), more gender stereotyped (Hypothesis 1b [H1b]), and feel more pressure for gender conformity (Hypothesis 1c [H1c]) than CE students. We also tested the exploratory hypothesis that SS students would report more exposure to gender equality than CE students (Hypothesis 1d [H1d]). This hypothesis is based on the observation that SS schools often have students from families that are wealthier, and these schools are more Western oriented and sometimes explicitly or historically communicate egalitarian missions and values (Liben, 2015; W. Y. Ho, 2004). SS girls' school stakeholders including the students, parents, and teachers generally endorsed antigender discrimination as a rationale for SS schooling more strongly than did CE stakeholders (e.g., endorsing that girls were treated unfairly in the past, and so they deserve extra help to catch up with boys; Pahlke, Bigler, & Patterson, 2014).

H2: Longitudinal Change

After high school graduation, the SS students would enter the mixed-gender world and thereby encounter more opportunities to interact with other-gender peers than they had prior to graduation. Based on gender socialization and intergroup contact theories (Bussey & Bandura, 1999; Cook et al., 2022; Mehta & Strough, 2009; Pettigrew & Tropp, 2006), we hypothesized that any differences between school types would therefore diminish after high school graduation (H2). Support for this hypothesis would be demonstrated if there were statistically significant paths from school type to the latent change variables in a latent change score (LCS) model (Coman et al., 2013; Kievit et al., 2018; see the Main Analyses section for additional explanation).

H3: Cross-Lagged Effects

We used a CLP model to explore potential mediation associations among the study variables (Orth et al., 2021). Specifically, we examined the predictive associations between school type and the gender cognition variables, as well as among the gender cognition variables themselves (H3). We first tested the significance of the paths from school type to the gender cognition variables at initial testing (Time 1 [T1]) and then tested whether these baseline effects passed on to later scores (on either the same or different outcomes at Time 2 [T2]). We explored all potential effects, but the effects of gender salience were of particular interest due to the centrality of this construct in the SS schooling debate.

Method

Participants

Recruitment

Hong Kong students in Form 6 (i.e., final year of high school; $M_{\text{age}} = 17.63$) were invited to complete a Qualtrics questionnaire online before (T1) and after (T2) high school graduation. Given the predominance of CE schools, the general reluctance of schools to take part in research, the local sociopolitical unrest in 2019–2020, and the requirement of follow-up testing during a life transition, recruiting participants for this study was challenging. To

maximize the participant pool and increase sample diversity, we recruited from any channel that would allow us to reach Form 6 students rather than recruiting from only a few specific schools. Indeed, judging from the students' reports on the schools they were attending, the sample of participants came from over 100 different schools. On average, no school was represented by more than a few students. This removes concerns about clustering or skewing of the results by a few schools. Initial invitations at T1 were distributed through local high schools, online forums, and chat groups. Participants were encouraged to advertise the study to other eligible students. Potential participants were told that this was a longitudinal study on students' cognitive and social development before and after graduation; the recruitment materials did not emphasize the focus on gender cognitions, thus minimizing potential self-selection bias based on students' gender cognitions. They were informed if they agreed to participate, they would be invited to complete an online questionnaire (about 45 min) twice, once in their final year of high school and again about a year after graduation. They were explicitly informed that they could decide whether to participate again when contacted the second time. After completing each of the two assessment questionnaires, participants received a coupon worth about USD13 and were eligible to join an online support group formed by our research assistants for them to consult about future study and career plans.

Sample Size

No clear rule exists regarding the proper sample size for the types of analyses adopted in this research (propensity score-matched, cross-lagged data in structural equation modeling), and much depends on unknown data parameters (e.g., actual variances). However, prior studies collecting new empirical data on SS and CE students typically had groups of fewer than a hundred participants to several hundred (Pahlke, Hyde, & Allison, 2014). Given that the general rule of thumb for structural equation modeling is a total sample of 200 (Boomsma, 1985; Kline, 2016), we judged that a 400-student sample, roughly equally split between SS and CE schools, would be adequate. A total of 700 students submitted the T1 questionnaire between September 2019 and April 2020. After excluding those respondents who failed all three attention check questions which asked participants to select a preassigned option ($n = 28$), or who reported a nonbinary gender identity ($n = 5$), the final T1 sample included 667 participants (437 CE students, 230 SS students, 381 self-reported as female, 286 as male, $M_{\text{age}} = 17.63$ years, $SD = 0.72$). Because in Chinese, the terms gender and sex are not distinguished and no single term worked for all, we refer to participants as male or female students.

T1 participants were contacted again after graduation to complete the T2 questionnaire. A total of 497 participants responded to the T2 questionnaire between January 2021 and August 2021 (attrition rate = 25.5%). After excluding the respondents who completed fewer than half of the survey questions ($n = 25$), failing all three attention-check questions ($n = 7$), providing ambiguous information on school type ($n = 1$), or repeating high school ($n = 1$), the final sample at T2 consisted of 463 participants who were included in further analyses (165 CE students self-reported as female, 121 SS students self-reported as female, 135 CE students self-reported as male, 42 SS students self-reported as male, $M_{\text{age}} = 19.01$ years, $SD = 0.75$; average time interval of response in months = 16.64, $SD = 2.00$). Table 1 shows detailed demographic characteristics of the final sample. The demographic

Table 1
Descriptive Statistics of Demographic Variables in the Final Sample

Variable	All						Female						Male					
	Total sample		CE		SS		CE		SS		CE		SS		CE		SS	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age at T1 (range = 15.80–22.96)	17.62	0.74	17.66	0.81	17.55	0.58	17.65	0.80	17.52	0.62	17.67	0.82	17.64	0.47	17.67	0.82	17.64	0.47
Age at T2 (range = 17.42–24.08)	19.01	0.75	19.04	0.81	18.95	0.60	19.07	0.82	18.94	0.64	19.01	0.81	18.97	0.47	19.01	0.81	18.97	0.47
Income ^a (range = 1–7)	3.25	1.33	3.00	1.18	3.70	1.46	2.87	1.05	3.72	1.52	3.16	1.32	3.64	1.28	3.16	1.32	3.64	1.28
Father education ^b (range = 1–6)	3.82	1.12	3.61	1.01	4.19	1.21	3.57	0.94	4.17	1.20	3.66	1.08	4.24	1.24	3.66	1.08	4.24	1.24
Mother education ^b (range = 1–6)	3.78	1.09	3.59	1.06	4.15	1.05	3.54	0.97	4.17	1.09	3.64	1.17	4.08	0.92	3.64	1.17	4.08	0.92
School banding (range = 1–3)	1.42	0.68	1.51	0.76	1.25	0.46	1.53	0.80	1.26	0.44	1.50	0.71	1.19	0.51	1.50	0.71	1.19	0.51
Number of brothers (range = 0–5)	0.48	0.60	0.51	0.63	0.44	0.56	0.56	0.59	0.40	0.55	0.44	0.66	0.55	0.55	0.44	0.66	0.55	0.55
Number of sisters (range = 0–5)	0.56	0.73	0.62	0.79	0.43	0.59	0.69	0.85	0.45	0.55	0.54	0.70	0.38	0.70	0.54	0.70	0.38	0.70
Religiosity level ^c (range = 1–7)	2.53	1.93	2.56	1.95	2.47	1.91	2.53	1.92	2.64	1.97	2.59	1.98	2.00	1.67	2.59	1.98	2.00	1.67
N	463		300		163		165		121		135		42		135		42	

Note. Results are based on raw data without imputation or matching. CE = coeducational school; SS = single-sex school; T1 = Time 1 measure; T2 = Time 2 measure.

^a Response scale for monthly family income (in HKD): 1 = less than 10,000; 2 = 10,000–20,000; 3 = 20,000–40,000; 4 = 40,000–60,000; 5 = 60,000–100,000; 6 = 100,000–150,000; 7 = more than 150,000. ^b Response scale for father education and mother education: 1 = no schooling/preprimary; 2 = primary; 3 = junior secondary; 4 = senior secondary; 5 = postsecondary (diploma/certificate, subdegree course, degree course); 6 = postgraduate. ^c Response scale for religiosity level: 1 = not at all; 2 = only nominal; 3 = very weak; 4 = weak; 5 = moderate; 6 = strong; 7 = very strong.

comparison between participants retained and participants who dropped out at T2 is reported in Table S1 in the online supplemental materials. Overall, there was no significant difference between the retained and attrited groups on the demographic variables.

Ethics approval was obtained from the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong. Participants provided informed consent and completed measures of demographic variables (school type, gender, birth date, parent education level, family income, school banding [an indicator of school quality], religion, ethnicity) at T1 and a battery of measures on gender cognitions (gender salience, gender stereotypes, felt pressure for gender conformity, exposure to gender equality) and interpersonal associations at both T1 and T2. In this study, we focus on the gender cognitions, so only this set of measures and results are reported. The Chinese versions were translated by Sylvia Yun Shi and Wang Ivy Wong and a research assistant with a professional degree in translation.

Measures

Gender Salience

Gender salience has been tapped in various ways (e.g., Cota & Dion, 1986; McGuire et al., 1979; Mehta, 2015; Palomares, 2008; Serbin & Sprafkin, 1986; Tobin et al., 2010). We focus on measures that assess the accessibility and relevance of gender in schemas, manifested as spontaneously noticing gender and making gender-based categorizations. We measured gender salience about both one's own gender and others' gender, using McGuire et al.'s (1979) open-ended probes and their modification. To assess the salience of one's own gender, participants were asked to give three short answers to each of these two questions: "Tell me what you are" and "Tell me what you are not" (McGuire et al., 1979). This measure detected individual and group differences in gender salience even when subjects were tested outside the environment producing those differences (McGuire et al., 1979) and were used successfully in a prior cross-sectional study of SS and CE high school and college students (Wong et al., 2018). It also better reflects gender salience than do related measures of gender identity (Cota & Dion, 1986). Salience about others' gender was measured by a modified measure derived from this method. Participants were given the instruction "Imagine you met another student on the first day of college (for T1) or coworker on the first day of work (for T2) and you start talking," and then asked to describe "This person is ..." and "This person is not ..." each with three short phrases, similar to the measure for gender salience about self. For both self and others' gender salience measures, the answers were scored as "1" if they included gender or gender-related terms (e.g., "I am a girl") or "0" if they made no reference to gender (e.g., "I am a student"), with a maximum score of 6 each. Coding was first done by two researchers separately and the disagreements were then resolved by discussion. The interrater reliabilities prior to resolving the disagreements were good ($k_s \geq .80$ at T1 and $\geq .90$ at T2). Item scores of each measure were summed, with higher sum scores meaning higher gender salience.

Gender Stereotypes

The Social Roles Questionnaire (Baber & Tucker, 2006) was used to measure students' endorsement of gender stereotypes. This measure was chosen because it has shown construct validity and does not

correlate with social desirability (Baber & Tucker, 2006). It consists of five gender-transcendent items (e.g., "People can be both aggressive and nurturing regardless of sex") and eight gender-linked items (e.g., "A father's major responsibility is to provide financially for his children"). Participants responded on a percentage scale ranging from 0% = *strongly disagree* to 100% = *strongly agree* with increments of 10%. The five gender-transcendent items were reverse coded. Item scores were averaged, and higher mean scores indicate stronger gender-stereotyped beliefs (Cronbach's $\alpha_s = .82$ for T1 and $.86$ for T2).

Felt Pressure for Gender Conformity

The felt pressure scale was originally designed by Egan and Perry (2001) for adolescents and then revised for children by asking how others might respond if the participant were to do something that others of their gender usually do not do (see Patterson, 2012 and Schroeder & Liben, 2021). We further adapted 12 items to measure felt pressure for gender conformity in the adolescent/adult sample. Specifically, participants were asked to rate how much they think they would be pressured if they wanted to do something that men/women usually do (the gender is not the same as that of the participant's own). Pressure from four sources of people (parents/friends/classmates/other relatives) was rated separately. First, they rated how much these people would try to stop them (four items) and then how much they think these people would tease them (four items). Finally, they rated how much they think these people would try to get them to act more like women/men (the gender is the same as the participant's own; four items). Responses were on a 4-point scale (1 = *not at all* to 4 = *a lot*). Item scores were averaged, and higher mean scores indicate stronger felt pressure for gender conformity (Cronbach's $\alpha_s = .85$ for T1 and $.87$ for T2).

Exposure to Gender Equality

The Exposure to Feminism Scale developed in a study of adolescent girls (Leaper & Brown, 2008) was adapted by changing the focus from feminism to gender equality. Participants were given a definition of gender equality in a format similar to that used in the original study. In particular, participants were told:

Gender equality refers to the belief in equality for women and men. People who believe in gender equality point to ways that society and some people treat boys and men or girls and women in unfair ways. For example, sometimes women are turned down for jobs or boys are not allowed to play with certain toys—because they are female or male. Another example is when women or men get unwanted sexual comments. When discrimination like this occurs, it is known as sexism.

After having read the definition of gender equality, participants were asked to indicate if they had learned about gender equality from each of six different sources (literature, media, parents, other family members, teachers, peers) by choosing *yes* (coded as 1) or *no* (coded as 0). Item scores were summed, and higher sum scores indicate more sources of exposure to gender equality.

Data Analyses

Multiple Imputation

For all study variables, missing data were <8.86% at T1 and <0.44% at T2. Little's missing completely at random test

suggested that the missingness was completely at random, $\chi^2(635) = 62.34, p > .999$. Following the best-practice recommendations for handling missing data, we employed multiple imputation (Graham, 2009) using the Multivariate Imputation by Chained Equations (mice) package (Version 3.15.0) in R (van Buuren & Groothuis-Oudshoorn, 2011). To our knowledge, there is no strict rule regarding the number of data sets that should be imputed. The rule of thumb is that the number of imputed data sets should be at least as large as the percentage of incomplete cases (White et al., 2011). We imputed 30 data sets which exceeded the minimum required number of imputed data sets for the missingness in our data. Auxiliary variables correlated with the imputed variables at $r > .40$ level, which would be considered as a medium to large association according to the operational convention (Cohen, 2013), were included in the predictor matrix of the imputation model.

Propensity Score Matching

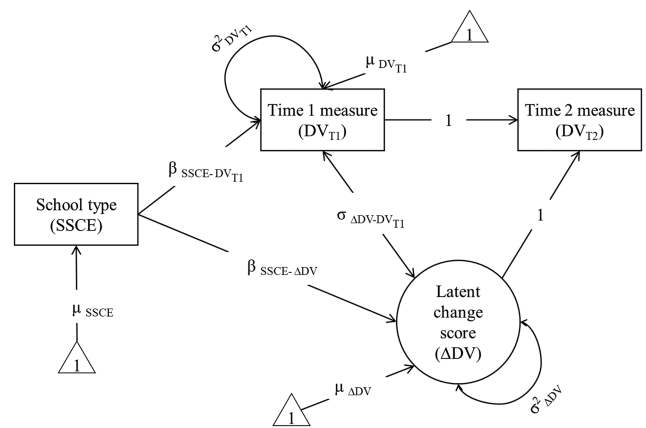
To approximate our observational data to experimental data and control for potential confounding factors, we implemented propensity score matching (Nagengast et al., 2013; Rosenbaum & Rubin, 1983) on the imputed data sets using the MatchThem (Version 1.0.1) and MatchIt (Version 4.5.2) packages in R (D. E. Ho et al., 2011; Pishgar et al., 2021). The procedure of propensity score matching mainly consists of two stages (Stuart, 2010). In the design stage, the covariates and the matching method are selected and the quality of the matched samples is assessed. In the analysis stage, the analysis of the outcome variables is conducted using the well-matched samples. For multiple imputed data sets, matching and analysis are done on each of the imputed data sets and the results are pooled together using Rubin's rules (Pishgar et al., 2021; Rubin, 1987).

In the design stage, we adopted a conservative criterion to include the demographic variables that correlated with at least one of the outcome variables at $p < .20$ level as the covariates for matching (Cheslack-Postava et al., 2015). Based on this criterion, eight covariates were included in matching (i.e., age, family income, father education, mother education, school banding, number of brothers, number of sisters, religiosity level). We then applied optimal full matching to estimate the average effect of treatment on the treated. Optimal full matching is potentially the best matching technique (Hansen, 2004). It uses all available subjects in the treatment and the control groups (i.e., the SS and CE groups in this study), places comparable participants into subgroups of different sizes to compute the propensity score weights, and optimizes the size of each subgroup to automatically maximize the overall achieved balance (Hansen, 2004; Stuart, 2010). The average effect of treatment on the treated in the current study is the effect of SS schooling for the population of students who attended an SS school or have the same characteristics as those who attended an SS school. Diagnostics of covariate balance suggested that optimal full matching yielded an excellent balance between the SS and the CE groups. The absolute standardized mean difference of all covariates after matching was well below the cutoff of 0.25 (Harder et al., 2010). Comparisons of the achieved balance also suggest that optimal full matching performed better in our data than other common matching approaches (i.e., nearest neighbor matching and optimal pair matching). The detailed balance measures are shown in Table S2 in the online supplemental materials.

Main Analyses

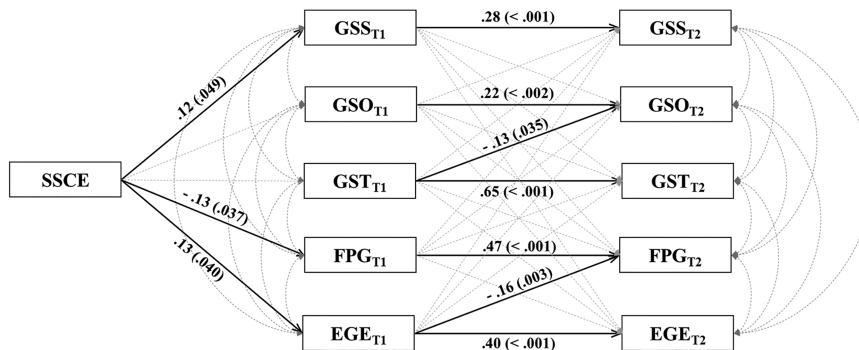
The main analyses were conducted using the lavaan.survey (Version 1.1.3.1) package in R (Oberski, 2014) and were conducted using the matched data sets after multiple imputation. First, we tested the zero-order correlations among all study variables as preliminary analyses. Second, we used LCS models (Coman et al., 2013; Kievit et al., 2018) to investigate whether gender cognitions differed by school type, whether they changed from T1 to T2, and whether these changes (if any) varied by school type or gender (see Figure 1 for the general path diagram). Each gender cognition outcome was tested in a separate model. Finally, we explored the potential mediation effects of the gender cognition outcomes in a CLP model (Orth et al., 2021; see Figure 2 for the path diagram). Note that the school-type variable (i.e., SS vs. CE) represented a status prior to T1. Therefore, the two-wave measure of the gender cognition variables, along with the school-type variable, provided sufficient grounds for testing the significance of the mediation paths. We first tested all possible paths from school-type to T1 variables (path *a*) and from T1 to T2 variables (path *b*), and then we tested the indirect effect of $a \times b$ when both *a* and *b* paths were significant. If the indirect effect of $a \times b$ was significant, we could conclude that there was a significant mediation effect (if path *b* was a cross-lagged path) or a significant autoregressive indirect effect (if path *b* was an autoregressive path). The significance of the indirect effect parameters was tested using the delta method combined with robust maximum likelihood estimation (see Oberski, 2014 for technical explanations). In this regard, although the LCS and CLP both addressed school-type differences, they also addressed different questions such as whether longitudinal changes differed by school type versus whether gender salience mediated between school-type and other gender cognition variables. Sensitivity analysis was conducted by also using the raw data set without imputation and without matching (see Tables S5–S7 in the online supplemental materials). All *p* values reported are two-sided.

Figure 1
Path Diagram of the LCS Models



Note. The residual error and the intercept of DV_{T2} were set to zero. The autoregressive path and the ΔDV to DV_{T2} path were set to 1. LCS = latent change score; DV = dependent variable; T1 = Time 1 measure; T2 = Time 2 measure; SSCE = school type (0 = coeducational school, 1 = single-sex school).

Figure 2
Path Diagram of the CLP Model



Note. Significant regression, autoregression, and cross-lagged paths ($p < .05$) are marked with solid-line arrows. Values beside each significant path are standardized regression coefficients β (outside the parentheses) and p values (inside the parentheses). Nonsignificant regression estimates and variance-covariance estimates are omitted from this graph for simplicity. The mediation effect of exposure to gender equality between school type and felt pressure, the autoregressive indirect effects of gender salience about self, exposure to gender equality and felt pressure were not significant, $ps > .05$. Full results can be found in Table S4 in the online supplemental materials. Results are based on the multiple imputed and matched data sets and pooled using Rubin's rules. CLP = cross-lagged panel; GSS = gender salience about self; T1 = Time 1 measure; T2 = Time 2 measure; GSO = gender salience about others; GST = gender stereotypes; SSCE = school type (0 = coeducational school, 1 = single-sex school); FPG = felt pressure for gender conformity; EGE = exposure to gender equality.

Transparency and Openness

We reported how we determined our sample size, data exclusion considerations, and the selection of measures in this study. This study was not preregistered. The data and measures are available from the corresponding author upon reasonable request. Data were analyzed using R, Version 4.2.2 (R Core Team, 2022) and packages mice, Version 3.15.0 (van Buuren & Groothuis-Oudshoorn, 2011), MatchIt, Version 4.5.2 (D. E. Ho et al., 2011), MatchThem, Version 1.0.1 (Pishgar et al., 2021), and lavaan.survey, Version 1.1.3.1 (Oberski, 2014).

Results

We report below the results organized by hypotheses and the type of analyses. Table 2a summarizes the hypothesis testing results and Table 2b summarizes meaningful exploratory findings.

Preliminary Analyses

Measurement invariance testing was conducted on the measures of gender stereotypes and felt pressure for gender conformity across groups and across time. Following common practice and recommendations in psychological research (Little, 2013), strong invariance is adequate for the interpretability of the results. Using the evaluation criterion of $\Delta CFI \leq .01$, both measures achieved strict invariance across SS versus CE groups. However, for invariance across gender groups, only the T2 felt pressure for gender conformity measure achieved strong invariance, while T1 felt pressure and the gender stereotypes measure at both time points achieved weak invariance. Regarding across-time invariance, the gender stereotypes measure achieved strict invariance and the felt pressure measure achieved strong invariance. Measurement invariance testing was not performed on the other measures

(gender salience and exposure to gender equality) because they were not psychological scales requiring such tests.

Table 3 shows descriptive statistics and zero-order correlations among key study variables at each time point on SS and CE students matched after propensity score matching and multiple imputation. Note that normative gender differences in the outcome variables resembling those reported in the Western literature (Cook et al., 2022; Perry et al., 2019) were found, suggesting that the measures performed in reliable ways and the data were meaningful. At both time points, male students reported higher gender stereotypes and felt more pressure for gender conformity than female students, whereas female students reported higher gender salience about self (but not about others) and more exposure to gender equality than male students.

To test the associations between school type and gender cognitions and the change over time, we used LCS models because these models could investigate school-type differences, changes over time, and the associations of the changes with school type simultaneously. To examine the predictive associations among different variables, we used the CLP model. Both models are described below. Multigroup analyses additionally addressed whether the LCS and CLP findings differed by student gender.

School-Type Differences (H1) and the Longitudinal Changes Surrounding High School Graduation (H2)

Table 4 shows the results of the LCS models (because the LCS models were saturated with $df = 0$, the model fit indices could not be estimated). School type significantly predicted gender salience about self, felt pressure for gender conformity, and exposure to gender equality at T1, but not gender salience about others or gender stereotypes. Before graduation, SS students reported higher gender salience about self ($M = 1.28$, $SD = 0.99$ for SS students; $M =$

Table 2
Summary of Key Findings

Research question	Hypothesis tested	Extent of support
(a) Summary of hypothesis testing results		
School-type differences	H1a: SS school students would be more gender salient. H1b: SS school students would be more gender stereotyped. H1c: SS school students would feel more pressure for gender conformity. H1d: SS students would report more exposure to gender equality than CE students.	Supported for H1a for gender salience about self; not supported for gender salience about others; Not significant for H1b gender stereotypes; Contrary to hypothesis for H1c felt pressure; Supported for H1d exposure to gender equality. (Indicated by corresponding regression paths between SSCE and gender cognitions at T1 in the LCS models.)
Longitudinal differences	H2: School-type differences would reduce after high school graduation.	Not supported. (Indicated by nonsignificant paths from school type to the latent change score variable in the LCS models.)
Mediation effects	H3: School type would have indirect effects on later scores (either on the same or different outcomes). The mediation effects of gender salience were of particular interest.	Not supported. (Indicated by the lack of significant indirect effects in the CLP model.)
(b) Summary of meaningful exploratory findings		
Gender differences	Normative gender differences were found at both time points. Male students reported higher stereotypes and felt pressure for gender conformity; female students reported higher gender salience about self and more exposure to gender equality. (Indicated by zero-order correlations.)	
Moderation by gender	Differences between school types and variable relationship paths did not differ significantly by student gender. (Indicated by multigroup analyses of the LCS models and the CLP model.)	
Role of exposure to gender equality	Higher exposure to gender equality at T1 predicted lower felt pressure at T2. (Indicated by the CLP model.)	
Role of gender stereotypes	Higher gender stereotypes at T1 predicted less salience about others' gender at T2. (Indicated by the CLP model.)	

Note. H = hypothesis; SS = single-sex school; CE = coeducational school; SSCE = school type (single-sex or coeducational); LCS = latent change score; CLP = cross-lagged panel.

1.03, $SD = 1.00$ for CE students), felt less pressure for gender conformity ($M = 2.18$, $SD = 0.55$ for SS students; $M = 2.34$, $SD = 0.60$ for CE students), and had more exposure to gender equality ($M = 4.34$, $SD = 1.26$ for SS students; $M = 4.00$, $SD = 1.28$ for CE students) than CE students. The effects sizes were small (β ranged from .12 to $-.13$). The LCS models additionally showed that there was a significant main effect of time on gender salience about self (but not about others), felt pressure for gender conformity, and gender stereotypes. However, school type did not significantly predict the changes over time. Students in general reported an increase in gender salience about self, a decrease in felt pressure for gender conformity, and a decrease in gender stereotypes from T1 to T2, regardless of which type of school they attended. Although the zero-order correlations (Table 3) showed that the school-type differences in some outcomes changed from statistically significant at T1 to nonsignificant at T2 (gender salience about self, felt pressure for gender conformity, and exposure to gender equality), the magnitudes of these differences were similar for the two time points. Coupled with the LCS model, which showed no significant association between school type and the change over time, the overall findings provide no strong evidence to support H2 that school-type differences were significantly diminished from T1 to T2.

To test whether the school-type and time differences were generalizable across male and female students, multigroup analysis by gender was conducted. First, the intercept of LCS was fixed to examine whether the time effect differed by gender. Second, the regression path using school type to predict each T1 outcome was fixed to examine whether gender moderated the school-type differences in T1 outcomes. Third, the path from school type to LCS was fixed to examine whether gender moderated the association between school type and changes over time. The likelihood-ratio test on the gender salience about others was significant when the regression path from school

type to T1 variable was fixed ($p = .044$), indicating that there was a school type by gender interaction on T1 gender salience about others. However, follow-up analyses did not show significant school-type differences for either gender (male SS students reported lower gender salience about others than male CE students, whereas female SS students reported higher gender salience about others than female CE students, but neither difference was significant: for male students, $\beta = -.12$, $p = .165$; for female students, $\beta = .13$, $p = .137$). Results of multigroup analysis on the other outcomes also failed to reveal significant effects (Table 4), meaning that no significant gender differences were found on the school-type differences in T1 outcomes or the time effects for gender salience about self, gender stereotypes, felt pressure, and exposure to gender equality.

Cross-Lagged Associations Among the Outcomes (H3)

Figure 2 shows the results of the CLP model (for detailed statistics, see Table S4 in the online supplemental materials). The fit indices indicated that this model fitted well on our data (comparative fit index [CFI] = 1.000, root-mean-square error of approximation [RMSEA] = .000, standardized root-mean-square residual [SRMR] = .009). As with the LCS analyses, at T1, school type significantly predicted gender salience about self, felt pressure for gender conformity, and exposure to gender equality (β ranged between .12 and $-.13$). SS students were more salient about their own gender, reported more exposure to gender equality, and felt less pressure for gender conformity than CE students. The novel findings from the CLP models were those concerning the autoregressive paths, cross-lagged paths, and mediations. The autoregressive paths for all five outcome variables were significant, with T1 outcomes positively predicting T2 outcomes, suggesting expected stability in relative scores in these variables (those with relatively higher scores at T1

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Zero-Order Correlations of the Study Variables

Variable	Observed range	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. School type ^a	NA	0.35	0.48	—										
2. Gender ^b	NA	1.59	0.49	.23***	—									
3. Gender salience about self (T1)	0-5	1.12	1.00	.12*	.22**	—								
4. Gender salience about self (T2)	0-5	1.28	1.04	.05	.22***	.30***	—							
5. Gender salience about others (T1)	0-5	0.68	0.94	-.01	.01	.24***	.09	—						
6. Gender salience about others (T2)	0-5	0.65	0.95	-.01	.09	.03	.22***	.21**	—					
7. Gender stereotypes (T1)	0-6.62	3.15	1.36	-.07	-.35***	.02	-.07	.05	-.11	—				
8. Gender stereotypes (T2)	0-7.46	2.90	1.41	-.10	-.28***	.01	-.02	.07	.00	.65***	—			
9. Felt pressure for gender conformity (T1)	1-4	2.28	0.59	-.13*	-.38***	-.07	-.14*	-.02	-.03	.23**	.15*	—		
10. Felt pressure for gender conformity (T2)	1-4	2.17	0.59	-.12	-.32***	-.09	-.06	.01	.13	.16	.13	.51***	—	
11. Exposure to gender equality (T1)	1-6	4.12	1.29	.13*	.19*	.10	.06	.10	-.01	-.11	-.08	-.16	-.23**	—
12. Exposure to gender equality (T2)	0-6	4.00	1.37	.12	.22**	.13	.09	.04	-.06	-.13	-.09	-.12	-.21*	.42***

Note. The reported means, standard deviations, and correlation coefficients are based on the multiple imputed and matched data sets using Rubin's rule. The possible ranges of the variables are 0-6 for gender salience about self and about others, 0-10 for gender stereotypes, 1-4 for felt pressure for gender conformity, and 0-6 for exposure to gender equality. NA = not applicable; T1 = Time 1 measure; T2 = Time 2 measure.

^a 1 = single-sex school and 0 = coeducational school. ^b 1 = male and 2 = female.

* $p < .05$. ** $p < .01$. *** $p < .001$.

tended to have relatively higher scores at T2). Besides, a significant cross-lagged path emerged as exposure to gender equality at T1 negatively predicted felt pressure for gender conformity at T2, meaning students who had more exposure to gender equality before graduation felt less pressure for gender conformity after they graduated. Another significant cross-lagged path was that gender stereotypes at T1 negatively predicted gender salience about others at T2, suggesting that students who had higher gender stereotypes before graduation had less gender salience about others after graduation.

We tested the mediation effect of exposure to gender equality between school type and felt pressure, and the autoregressive indirect effects of gender salience about self, exposure to gender equality, and felt pressure (i.e., indirect effect of $a \times b$ was tested when both a and b paths were significant). None of the mediation or autoregressive indirect effects was significant, $ps > .05$ (for detailed statistics, see Table S4 in the online supplemental materials).

Multigroup analysis by gender was conducted by fixing all the regression and covariance paths to be the same between the two gender groups. The likelihood-ratio test result was not significant, $p > .05$, thus providing no evidence that gender was a significant moderator of the associations found in the CLP model.

Discussion

In many parts of the world, gender-segregated schools represent a significant portion of the education system, and scholars and practitioners have engaged in much debate about the consequences and values of SS schooling (e.g., Brown, 2013; Halpern et al., 2011; Liben, 2015; Pahlke & Hyde, 2016; Sax, 2011; Signorella & Bigler, 2013). Recent research has suggested SS-CE differences in academic outcomes were largely effects of selection bias (Pahlke, Hyde, & Allison, 2014; Signorella & Bigler, 2013; Signorella et al., 2013). Another heatedly debated area about the consequences of gender-segregated schooling is whether SS schooling strengthens or reduces gender cognitions. Citing factors related to gender ratios within groups, the use of gender-differentiated language, and the assignment of gender-differentiated roles or activities, supporters of SS schooling reasoned that SS schooling would weaken gender cognitions while critics of SS schooling predicted the opposite.

Both of these contrasting predictions are compatible with the classic social-constructivist theories related to SS schooling, such as DIT (Bigler, 1995; Bigler & Liben, 2006, 2007; Liben, 2017) and socialization theories (Bussey & Bandura, 1999; Cook et al., 2022; McHale et al., 2003; Pettigrew & Tropp, 2006). Complex outcomes are possible because multiple forces are operating in natural environments. Characteristics of SS and CE schools may drive differences in contrasting directions. The theoretical debate about how SS and CE students differ renders direct empirical evidence even more important.

In an attempt to expand the corpus of relevant research, we conducted a longitudinal cross-lagged study that focused on several new or understudied gender cognition constructs, including gender salience, gender stereotypes, felt pressure for gender conformity, and exposure to gender equality. Our work included both male and female students who were enrolled in either SS or CE schools, and we collected data two times (the first during the final year of high school and the second about a year and a half later), thus bridging the critical transition marked by high school graduation. In addition to using a longitudinal cross-lagged design, we applied propensity score matching to compensate for the fact that it is impossible to

Table 4
Results of the LCS Models

Dependent variable of each model	Path	Estimate	SE	CI (lower)	CI (upper)	β	p	Likelihood-ratio test for gender difference		
								$\Delta\chi^2$	Δdf	p
Model 1 GSS	GSS _{T1} ~ SSCE	0.25	0.13	0.00	0.49	.12	.049	0.54	1	.463
	Δ GSS ~ SSCE	-0.13	0.15	-0.43	0.17	-.05	.385	0.25	1	.619
	Δ GSS intercept	0.21	0.10	0.01	0.40	.17	.036	<0.01	1	.982
Model 2 GSO	GSO _{T1} ~ SSCE	-0.02	0.12	-0.26	0.21	-.01	.839	4.07	1	.044
	Δ GSO ~ SSCE	0.01	0.16	-0.31	0.33	.00	.954	0.91	1	.339
	Δ GSO intercept	-0.03	0.11	-0.25	0.18	-.03	.755	0.10	1	.751
Model 3 GST	GST _{T1} ~ SSCE	-0.20	0.18	-0.55	0.15	-.07	.253	1.63	1	.201
	Δ GST ~ SSCE	-0.10	0.15	-0.39	0.19	-.04	.493	0.12	1	.733
	Δ GST intercept	-0.22	0.10	-0.40	-0.03	-.19	.022	1.42	1	.233
Model 4 FPG	FPG _{T1} ~ SSCE	-0.17	0.08	-0.32	-0.01	-.13	.037	<0.01	1	.970
	Δ FPG ~ SSCE	0.02	0.07	-0.12	0.16	.02	.793	0.18	1	.669
	Δ FPG intercept	-0.12	0.05	-0.21	-0.03	-.20	.011	0.89	1	.346
Model 5 EGE	EGE _{T1} ~ SSCE	0.34	0.17	0.02	0.67	.13	.040	0.71	1	.398
	Δ EGE ~ SSCE	-0.01	0.18	-0.37	0.36	.00	.975	1.12	1	.290
	Δ EGE intercept	-0.12	0.12	-0.35	0.12	-.08	.331	0.99	1	.320

Note. Figure 1 illustrates the general path diagram of the LCS models. Full results can be found in Table S3 in the online supplemental materials. Because the LCS models were saturated with $df = 0$, the model fit indices could not be estimated. Results are based on the multiple imputed and matched data sets and pooled using Rubin's rules. Significant p values are bolded. A positive estimate of a $DV_{T1} \sim SSCE$ path indicates that the average value of the DV_{T1} is higher in SS students than in CE students. $\Delta = LCS (T2-T1)$; a positive ΔDV intercept indicates an increase in the DV from T1 to T2 averaged across SS and CE students. When ΔDV intercept is positive, a positive estimate of a $\Delta DV \sim SSCE$ path indicates a larger increase in the DV from T1 to T2 among SS students than among CE students; when ΔDV intercept is negative, a negative estimate of a $\Delta DV \sim SSCE$ path indicates a larger decrease in the DV from T1 to T2 among SS students than among CE students. For simplicity, estimates of the DV_{T1} intercept, DV_{T1} variance, and the covariance between DV_{T1} and ΔDV are not reported in this table. LCS = latent change score; CI = confidence interval; GSS = gender salience about self; T1 = Time 1 measure; SSCE = school type (0 = coeducational school, 1 = single-sex school); GSO = gender salience about others; GST = gender stereotypes; FPG = felt pressure for gender conformity; EGE = exposure to gender equality; DV = dependent variable; SS = single-sex school; CE = coeducational school; T2 = Time 2 measure.

randomly assign students to SS versus CE schools intentionally for research purposes. This approach allowed us to control for differences among students that would otherwise undermine the value of comparisons between those from SS and CE schools. To our knowledge, this study presents one of the most stringently designed data sets and analytical approaches to address several gendered outcomes that have not been examined together in earlier work.

Differences by School Type

Did SS Students Have Higher Gender Salience?

Our data showed that in the final year of high school, SS students were more gender salient than CE students, even when carefully matched on potentially confounding background variables which were correlated with key study variables. The data were consistent with arguments that have been made by earlier critics of SS schooling (Halpern et al., 2011; Liben, 2015; Pahlke, Hyde, & Allison, 2014; Signorella & Bigler, 2013; Signorella et al., 2013). SS students were indeed more gender salient about themselves, albeit only modestly. This finding is consistent with the prediction made by critics of SS schooling who draw on DIT (Halpern et al., 2011; Liben, 2015; Pahlke, Hyde, & Allison, 2014) and with prior cross-sectional findings reporting the same difference in high school students (Wong et al., 2018). We thus infer that SS schools may contain features that select for or increase the salience of gender in some ways, for example, by including gender in their school's name. Proponents of SS schools argue that they reduce gender salience in other ways, for example, by offering few opportunities to observe differential treatment of students in relation to gender. However, critics of SS education have provided counterarguments for this and other rationales for SS education

(e.g., see Bigler et al., 2014, in particular, their discussion of the rationale that SS education "works by reducing attention to gender"). These discussions underscore the importance of empirical data.

The data from the current study are consistent with the position that SS schools increase self-gender salience insofar as the data showed that students in SS schools were more likely than those in CE schools to include gender when they were asked to provide descriptions of themselves. Gender was thus a more accessible scheme of self-categorization among SS than CE students. At the same time, an association between school type and gender salience was not pervasive insofar as responses to the measure of gender salience about others were not related to school type and did not correlate strongly with gender salience about self ($r = .24$ at T1 and $r = .22$ at T2 in the zero-order correlations). Few other studies on spontaneous salience about others' gender are available, so whether the SS-CE difference in gender salience is restricted to their self-concept but not to their perception of others needs additional study.

Did SS Students Score Higher on Other Gender Cognition Outcomes?

Gender stereotypes assessed by the Social Role Questionnaire showed no significant difference by school type. In Pahlke, Hyde and Allison (2014) meta-analysis, gender stereotypes were significantly different by school type. Specifically, they were lower in SS girls in studies that included some control variables (i.e., higher quality studies), but the authors noted that this finding should be viewed with caution due to the heterogeneity of measures included as gender stereotype outcomes. Our failure to find any significant associations between school type and gender stereotyping is consistent with a recent finding by Lee and Sambanis (2023). These investigators

studied South Koreans in middle adulthood who had earlier been randomly assigned to attend either SS or CE schools. They found no SS versus CE differences on either the measure of hostile sexism or work-family gender role attitudes. Interestingly, in our work, we found that despite finding that SS students reported higher self-gender salience, SS students did not score higher on measures that may be perceived as reflecting stronger stereotyping. Specifically, they did not differ from CE students in gender stereotypes, and they reported feeling less pressure for gender conformity than CE students. The latter finding contrasts with one study in the United States reporting higher felt pressure in SS girls (Drury et al., 2013). However, our findings are in line with female SS graduates' subjective feeling that they felt "freedom to be themselves" (Jephcote, 2022) and with another study reporting that girls in SS classes were slower responding to feminine traits when asked whether these traits described them (Kessels & Hannover, 2008).

In addition, we assessed a novel construct—exposure to gender equality. On this measure, SS students more commonly reported having learned about gender equality from various sources than did CE students. This difference may be driven by school ethos stressing equality and the chance for students in SS schools to engage in both gender-normative and nonnormative tasks given the absence of other-gender peers to divide the labor. Stakeholders of SS schools hold stronger antigender discrimination belief (compared to CE stakeholders; Pahlke, Bigler, & Patterson, 2014), and modern SS schools were often initially established with an egalitarian vision (Liben, 2015; W. Y. Ho, 2004). Our finding of higher exposure to gender equality in SS schools may be reasonable if stakeholders of SS schools put their antigender discrimination attitudes into practice. In fact, it would be important to further study the exact contexts created and the practices conducted at SS and CE schools, in addition to studying school gender composition per se. For example, SS and CE schools may both instill or alleviate gendered concepts in students depending on how the teachers frame gender in their daily practices (Robinson et al., 2021). Different challenges may exist in SS and CE environments, and it may be unnecessary to solve potential problems of CE schooling by separating boys and girls into different schools and classes (Liben, 2015), or the reverse.

In sum, SS and CE students did not significantly differ on all dimensions of gender cognitions, and the significant differences (i.e., in levels of gender salience, felt pressure, and exposure to gender equality) were small, with the direction of those differences varying across constructs. The current findings thus neither imply uniform positivity nor negativity about associations between SS schooling and gender cognitions. Instead, they suggest that differences between SS and CE schooling may exist but are complex, multifaceted, and cannot be assumed to be unidimensional.

Longitudinal Changes

Did SS–CE Differences Diminish After Graduation?

The hypothesis about the reduction of school-type differences across time was generally not supported. The LCS models, which considered data from T1 and T2 simultaneously, revealed that school type did not predict changes in any outcome over time. Put differently, the magnitude of differences between SS and CE students was not significantly different between T1 and T2. Thus, there was no strong

evidence that graduation reduced the differences in gender cognitions between SS and CE students even after SS students no longer remained in a single-sex school environment. However, it is important to note that the SS–CE differences were small to begin with, perhaps reducing room for school-type differences to narrow over time.

Were There Changes in Gender Cognitions Over Time Regardless of School Type?

The main effects in the LCS models showing a reduction of felt pressure and gender stereotypes over time are consistent with a prior study reporting a decrease in gender stereotypes over a year regardless of gender-segregation status (Signorella et al., 1996). The findings that gender salience increased between T1 and T2 even as felt pressure and gender stereotypes decreased are not necessarily at odds but may reflect general age-related changes that occur in emerging adulthood. For example, young adults may be exposed to more social situations that require attention to gender, such as dating, thus making gender a more salient feature in how they define themselves compared to when they were younger. Accumulated exposure to gender-egalitarian knowledge and experiences (Leaper & Brown, 2008) may also lead to reduced gender stereotypes and felt pressure for gender conformity while simultaneously increasing gender salience.

In sum, the findings suggest that SS schooling may still be associated with graduates' gender cognitions even after they have graduated. It is important to note that the second wave of data collection was conducted less than 2 years after graduation and thus the patterns of stability and change we found across time may not generalize to longer time spans or to other outcomes. However, the findings appear to be consistent with studies on interpersonal outcomes including anxiety about mixed-gender interactions (Wong et al., 2018) and marriage outcomes (Sullivan et al., 2012), which showed differences by school type years after graduation (e.g., lower divorce rate among male CE graduates). More well-controlled research that spans longer periods and includes a richer battery of understudied outcomes is needed to evaluate SS schooling comprehensively.

Cross-Lagged Path Analyses

Did Earlier Gender Cognitions Predict Later Outcomes and Did They Mediate SS–CE Differences?

To give preliminary insights into potential mechanisms through which SS schooling might be channeled into later outcomes, we conducted cross-lagged analyses with autoregressive and cross-lagged paths. We tested all possible effects but found few cross-lagged effects and no mediation effects. One cross-lagged effect was that students who had more exposure to gender equality before graduation felt less pressure for gender conformity after they graduated, which is consistent with the general understanding of an association between felt pressure and liberal gender attitudes (Cook et al., 2022). We found an unpredicted cross-lagged path showing that stronger gender stereotypes predicted less attention to others' gender (i.e., lower other-gender salience). This finding had not been studied previously, but it may suggest that highly gender-stereotyped people are less aware of gender issues (Rogers, 2020). More importantly, we did not find significant cross-lagged mediation effects of gender salience or exposure to gender equality despite finding that school type showed significant differences on both variables at T1.

As a result, the current data do not support the notion that SS students would evidence different gender cognitions following graduation when T1 scores were controlled and when SS and CE students' backgrounds were matched as closely as possible. However, some caveats are needed before concluding that there are no mediation effects. Specifically, because mediations were examined by testing the paths of indirect effects (e.g., indirect effect of $a \times b$ was tested when both a and b paths were significant), one possible interpretation is that there are no mediation effects when $a \times b$ is not significant. An alternative interpretation may be that the effect size of path a (i.e., the initial differences by school type) was small to begin with, thus making it necessary to have larger samples to detect small mediation effects. The causal mechanisms through which school type produced differences between SS and CE students remain to be identified.

Were Gender Cognition Differences Between SS and CE Students Consistent Across Participant Gender?

There is no strong theoretical or empirical basis to predict how the differences between alternative school types are modified by student gender. Girls were more heavily represented in prior research samples on SS schooling, and findings have been mixed regarding how boys and girls are affected (Pahlke, Hyde, & Allison, 2014). In the current study, we found no significant moderation by student gender for any of our analyses, suggesting that participant gender had little impact on the associations we examined. (We note that although the multigroup analysis was significant in the LCS model for gender salience about others, follow-up analyses showed no significant difference in school type in both male and female students.) The changes over time and the associations among different gender cognitions also were not significantly different for male and female students. Nevertheless, it would be wise to conduct additional studies to learn if our findings are replicated.

Implications for Future Research and Interventions

Our findings regarding gender salience and stereotypes lead us to call for more theoretical and empirical work to consider what it really means to find higher gender salience in SS students. In the current study, gender salience was defined as greater attention to or greater consciousness about gender and the tendency to make gender-based categorizations (Bigler & Liben, 2006, 2007; Cota & Dion, 1986; Liben, 2017; McGuire, 1984; McGuire et al., 1979; McGuire & Padawer-Singer, 1976; Mehta, 2015; Palomares, 2008; Serbin & Sprafkin, 1986; Wong et al., 2018). In our data, gender salience but not stereotypes significantly differed between SS and CE students, and these two measures did not significantly correlate or mediate. More work is needed to understand whether and under what conditions greater attention to gender affects stereotypes.

Some investigators have demonstrated that increasing attention to gender does indeed lead to stronger gender stereotypes. For example, Hilliard and Liben (2010) asked some teachers to use gendered language in their preschool classrooms and to use gender to organize activities. In just 2 weeks, children in those classes showed a significant increase in their gender stereotypes and avoided playing with children of another gender, an aversion they had not shown 2 weeks earlier. No similar changes occurred in the comparison classes in which teachers did not highlight gender unnecessarily (e.g., saying "Good morning, children" instead of "Good morning, boys and girls").

Because children may be prone to make naive extrapolations about gender whereas adolescents and adults are more cognitively advanced and familiar with gender equality concepts (Leaper & Brown, 2008), paying attention to gender may have different causes and impacts at different ages. For children, it may be especially important to teach ways to reject sexist messages explicitly as illustrated in a program for elementary school children (Lamb et al., 2009). However, even adults may need to be taught to be more critical when attending to gender given that well-meaning antisexist campaigns (e.g., to encourage women to pursue careers in science or to highlight women's disadvantage in science and math fields) may inadvertently be communicating and reinforcing gender stereotypes or discouraging people to apply to fields that they find especially difficult (Liben, 2016; Williams & Ceci, 2015).

In short, people who are highly attentive to gender (i.e., highly gender salient) may include different kinds of subgroups that have different gender stereotype profiles. Some high salience people may have strong gender stereotypes and some may have strong nonsexist, gender-egalitarian beliefs. Some of our findings are consistent with this possible interpretation. First, our work showed that SS students had higher self-gender salience while they also reported lower felt pressure and greater exposure to gender equality. Second, we found that students who had higher gender stereotypes were less attentive to others' gender (i.e., lower other-gender salience) later. The speculation that there are multiple profiles of people characterized by different combinations of gender salience and gender stereotypes could explain the lack of an overall correlation or mediation effect between gender salience and other outcomes. In fact, a study comparing Swedish gender-neutral schools and traditional schools found that children in a gender-neutral school were less gender-stereotyped but not less likely to automatically encode gendered information (Shutts et al., 2017), supporting our speculation that the association between gender salience and stereotyping may be multidimensional. Because higher exposure to gender equality was related to lower felt pressure regardless of school type, exposing students to gender equality information may be an effective pedagogical strategy for educators to optimize students' gender cognition outcomes (such as lowering felt pressure to conform to gender norms) at both SS and CE schools.

It is also worth noting that gender salience has sometimes been used to refer to or used interchangeably with other terms such as gender schematicity and gender accessibility. Additionally, gender salience has been measured in different ways (e.g., Cota & Dion, 1986; Coyle & Liben, 2016; Liben & Bigler, 2002, 2017; McGuire et al., 1979; Serbin & Sprafkin, 1986; Wong et al., 2018). It has been conceptualized both as a situational characteristic and as a person attribute (see Liben, 2017; Liben & Bigler, 2002, 2017). Some studies measured the hypothesized outcomes of gender salience (e.g., endorsement of gender stereotypes in Kessels & Hannover, 2008) and some manipulated the environmental factors of gender salience (e.g., Hilliard & Liben, 2010). In the current study, we defined gender salience as the accessibility or awareness of gender and separated it from measurements of its hypothesized environmental precedents (e.g., SS or CE environment) and hypothesized outcomes (e.g., gender stereotypes). It was conceptualized as a person attribute that shows individual differences but is influenced by situational factors such as SS or CE schooling environments. SS–CE schooling contexts are long term, similar to the way that families provide long-term environments that may foster chronic individual differences in gender salience (see Cota & Dion, 1986; McGuire et al., 1979; also see

Wood & Eagly, 2015). Further discussions about the definition and measurement of gender salience may drive the field forward and guide new research.

Limitations

This study offers several important contributions to the SS schooling literature, but we acknowledge a number of limitations. First, the sample size is reasonably large but, as in prior research (see Pahlke, Hyde, & Allison, 2014), males from SS schools were underrepresented. A larger sample may be needed to detect school by participant gender interactions and mediations. The comparisons across gender groups were also weakened by the lack of strong invariance for gender groups in two measures, specifically for the gender stereotype measure at both T1 and T2 and the felt pressure for gender conformity measure at T1. Thus, results involving comparing these two measures for male and female students should be regarded with caution.

Second, the findings are based exclusively on students from Hong Kong. Given that Hong Kong is a very industrialized and Westernized society with a long history of British rule and influence, its education system closely resembles those found in other Western cultures, especially those in Britain. While others have reported finding normative gender differences in gender development across cultures (Gibbons, 2000; Wong & VanderLaan, 2020), we also found normative gender differences in gender stereotypes and felt pressure that are in line with those reported in the West (Cook et al., 2022). Taken together, these observations suggest that the current findings are likely generalizable to other Westernized cultures. In comparison to studies that recruited students from very few schools (e.g., Drury et al., 2013; Signorella et al., 1996), our recruitment was not school based, and thus our samples of SS and CE students represented a wide range of schools. Our sampling method thus provided school diversity, but that diversity was within a single society. Different patterns of findings might have emerged if samples had been drawn from SS schools in more varied cultural contexts.

Third, although propensity score matching has the potential to minimize differences between groups (e.g., Nagengast et al., 2013), our matching did not control for gender cognitions prior to initial entry to SS or CE schools. Although self-selection into SS or CE schools on the basis of preexisting gender cognitions is possible, actual school allocation is determined by a myriad of factors such as academic merit, geographical proximity, and partial random assignment by lottery. These considerations may have helped to minimize the probability that students entered SS or CE schools based on preexisting gender cognitions. However, we cannot rule out this possibility.

Fourth, we focused on only one particular 1.5-year interval and context—a focus that made it feasible for us to recruit a large sample of SS and CE students during a critical transition. This high school graduation transition was selected because it marks the time that SS students leave a gender-segregated school environment and become fully immersed in a mixed-gender environment (all universities are CE in Hong Kong). Changes emerged within this short interval. For all student groups, over time we found increased gender salience, reduced gender stereotypes, and lowered felt pressure for gender conformity. However, the length of our longitudinal work does not permit us to know about changes that might occur in the more distant future. Even after immersion in institutional gender segregation has ended, graduates of SS schools may continue to favor same-gender friends and interactions (Wong et al., 2018), occupations, and

environments. Self-selected SS environments may reflect a distal effect of earlier segregation.

Conclusions

The evaluation of the consequences of SS versus CE schooling is far from complete, and the current study does not permit all questions about the differences between SS and CE schooling to be answered. However, our work demonstrates the utility of studying more varied gender cognition outcomes of SS versus CE schooling. SS students were slightly more attentive to their own gender, reported more exposure to gender equality, and felt less pressured to conform to gender norms both before and after graduation, while they did not differ from CE students in gender salience about others or in gender stereotypes. The range of our findings indicates that in real-life SS and CE schooling contexts, differences predicted by theories may not be as simple or substantial as those found in more tightly controlled experimental studies of gender segregation or identity salience because it is impossible to control for and manipulate all forces that may be at play.

Because the patterns of differences in gender cognitions between SS and CE students were not always significant, and the significant differences were small and varied in direction, it appears that gender cognition differences between of SS and CE schooling are neither as alarming nor as beneficial as suggested in the scholarly literature or the public press. This more balanced conclusion resonates with the conclusion that the most positive outcome of SS schooling touted by supporters of SS schools—academic achievement—is minimal at best once corrections are made for confounds between SS and CE samples (Pahlke, Hyde, & Allison, 2014; Signorella et al., 2013). Apart from focusing on achievement-related constructs, it would be valuable to assess a variety of constructs in order to understand the differences between SS and CE schooling more fully. Future research will be advanced if researchers can implement longitudinal designs and avoid (or statistically control for) initial differences between SS and CE samples.

References

- Agence, France-Presse. (2013, February 26). *China turns to all-boys classes as girls progress*. <https://www.scmp.com/news/china/article/1159110/china-turns-all-boys-classes-girls-progress#/>
- Baber, K. M., & Tucker, C. J. (2006). The Social Roles Questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles, 54*(7–8), 459–467. <https://doi.org/10.1007/s11199-006-9018-y>
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*(4), 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>
- Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development, 66*(4), 1072–1087. <https://doi.org/10.2307/1131799>
- Bigler, R. S., Hayes, A. R., & Liben, L. S. (2014). Analysis and evaluation of the rationales for single-sex schooling. In L. S. Liben, & R. S. Bigler (Eds.), *The role of gender in educational contexts and outcomes. Advances in child development and behavior* (Vol. 47, pp. 225–260). Elsevier.
- Bigler, R. S., & Leaper, C. (2015). Gendered language: Psychological principles, evolving practices, and inclusive policies. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 187–194. <https://doi.org/10.1177/2372732215600452>
- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child*

- development and behavior* (Vol. 34, pp. 39–89). Elsevier. [https://doi.org/10.1016/s0065-2407\(06\)80004-2](https://doi.org/10.1016/s0065-2407(06)80004-2)
- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16(3), 162–166. <https://doi.org/10.1111/j.1467-8721.2007.00496.x>
- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, 50(2), 229–242. <https://doi.org/10.1007/BF02294248>
- Brown, C. S. (2013). Legal issues surrounding single-sex schools in the U.S.: Trends, court cases, and conflicting laws. *Sex Roles*, 69(7–8), 356–362. <https://doi.org/10.1007/s11199-011-0001-x>
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106(4), 676–713. <https://doi.org/10.1037/0033-295X.106.4.676>
- Cheslack-Postava, K., Keyes, K. M., Lowe, S. R., & Koenen, K. C. (2015). Oral contraceptive use and psychiatric disorders in a nationally representative sample of women. *Archives of Women's Mental Health*, 18(1), 103–111. <https://doi.org/10.1007/s00737-014-0453-4>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Coman, E., Picho, K., McArdle, J., Villagra, V., Dierker, L., & Iordache, E. (2013). The paired t-test as a simple latent change score model. *Frontiers in Psychology*, 4, Article 738. <https://doi.org/10.3389/fpsyg.2013.00738>
- Cook, R. E., Martin, C. L., Nielson, M. G., & Xiao, S. X. (2022). Contemporary cognitive approaches to gender development: New schemas, new directions, and new conceptualizations of gender. In D. P. VanderLaan, & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research* (pp. 125–157). Springer. https://doi.org/10.1007/978-3-030-84273-4_5
- Cooper, S. (2010). *Single-sex education in the UK and a selection of other English-speaking countries* [Online]. National Coalition of Girls' Schools. <https://www.ncgs.org/Pdfs/Resources/singlesex%20education%20in%20the%20UK%20and%20other%20English%20speaking%20countries.pdf>
- Cota, A. A., & Dion, K. L. (1986). Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality and Social Psychology*, 50(4), 770–776. <https://doi.org/10.1037/0022-3514.50.4.770>
- Coyle, E. F., & Liben, L. S. (2016). Affecting girls' activity and job interests through play: The moderating roles of personal gender salience and game characteristics. *Child Development*, 87(2), 414–428. <https://doi.org/10.1111/cdev.12463>
- Drury, K., Bukowski, W. M., Velásquez, A. M., & Stella-Lopez, L. (2013). Victimization and gender identity in single-sex and mixed-sex schools: Examining contextual variations in pressure to conform to gender norms. *Sex Roles*, 69(7–8), 442–454. <https://doi.org/10.1007/s11199-012-0118-6>
- Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, 37(4), 451–463. <https://doi.org/10.1037/0012-1649.37.4.451>
- Fabes, R. A., Martin, C. L., Hanish, L. D., & DeLay, D. (2018). Gender integration in coeducational classrooms: Advancing educational research and practice. *School Psychology Quarterly*, 33(2), 182–190. <https://doi.org/10.1037/spq0000266>
- Fabes, R. A., Martin, C. L., Hanish, L. D., Galligan, K., & Pahlke, E. (2015). Gender-segregated schooling: A problem disguised as a solution. *Educational Policy*, 29(3), 431–447. <https://doi.org/10.1177/0895904813492382>
- García-Gracia, M., & Donoso Vázquez, T. (2016). Mixed schools versus single-sex schools: Are there differences in the academic results for boys and girls in Catalonia? *International Journal of Inclusive Education*, 20(2), 149–167. <https://doi.org/10.1080/13603116.2015.1079269>
- Gee, K. A., & Cho, R. M. (2014). The effects of single-sex versus coeducational schools on adolescent peer victimization and perpetration. *Journal of Adolescence*, 37(8), 1237–1251. <https://doi.org/10.1016/j.adolescence.2014.08.011>
- Gibbons, J. L. (2000). Gender development in cross-cultural perspective. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 389–415). Erlbaum.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., Liben, L. S., & Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science*, 333(6050), 1706–1707. <https://doi.org/10.1126/science.1205031>
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609–618. <https://doi.org/10.1198/016214504000000647>
- Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15(3), 234–249. <https://doi.org/10.1037/a0019623>
- Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development*, 81(6), 1787–1798. <https://doi.org/10.1111/j.1467-8624.2010.01510.x>
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>
- Ho, W. Y. (2004). *Schooling boys and girls: The development of single-sex and co-educational schools in Hong Kong* [CUHK Electronic Theses & Dissertations Collection]. The Chinese University of Hong Kong. <https://repository.lib.cuhk.edu.hk/en/item/cuhk-324554>
- Jephcote, L. (2022). *Young women's reflections on attending a private girls' school* [Unpublished doctoral thesis]. Cardiff University. <https://orca.cardiff.ac.uk/id/eprint/157899/2/1474086%20Lisa%20Jephcote%20-%20final%20submission.pdf>
- Kessels, U., & Hannover, B. (2008). When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology*, 78(2), 273–289. <https://doi.org/10.1348/000709907X215938>
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, 33, 99–117. <https://doi.org/10.1016/j.dcn.2017.11.007>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford Press.
- Lamb, L. M., Bigler, R. S., Liben, L. S., & Green, V. A. (2009). Teaching children to confront peers' sexist remarks: Implications for theories of gender development and educational practice. *Sex Roles*, 61(5–6), 361–382. <https://doi.org/10.1007/s11199-009-9634-4>
- Leaper, C. (2022). Origins and consequences of childhood gender segregation: Toward an integrative developmental systems model. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research* (pp. 159–205). Springer. https://doi.org/10.1007/978-3-030-84273-4_6
- Leaper, C., & Brown, C. S. (2008). Perceived experiences with sexism among adolescent girls. *Child Development*, 79(3), 685–704. <https://doi.org/10.1111/j.1467-8624.2008.01151.x>
- Lee, A., & Sambanis, N. (2023). Does school environment shape gender differences in leadership and participation? Evidence from a natural experiment in South Korea. *SSRN Electronic Journal*. Advance online publication. <https://doi.org/10.2139/ssrn.4321251>
- Lee, Y., & Nakazawa, N. (2022). Does single-sex schooling help or hurt labor market outcomes? Evidence from a natural experiment in South Korea.

- Journal of Public Economics*, 214, Article 104729. <https://doi.org/10.1016/j.jpubeco.2022.104729>
- Liben, L. S. (2014). The individual ↔ context nexus in developmental intergroup theory: Within and beyond the ivory tower. *Research in Human Development*, 11(4), 273–290. <https://doi.org/10.1080/15427609.2014.967048>
- Liben, L. S. (2015). Probability values and human values in evaluating single-sex education. *Sex Roles*, 72(9–10), 401–426. <https://doi.org/10.1007/s11199-014-0438-9>
- Liben, L. S. (2016). We've come a long way baby (But we're not there yet): Gender past, present, and future. *Child Development*, 87(1), 5–28. <https://doi.org/10.1111/cdev.12490>
- Liben, L. S. (2017). Gender development: A constructivist-ecological perspective. In N. Budwig, E. Turiel, & P. D. Zelazo (Eds.), *New perspectives on human development* (pp. 145–164). Cambridge University Press. <https://doi.org/10.1017/CBO9781316282755.010>
- Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, 67(2), vii–147. <https://doi.org/10.1111/1540-5834.t01-1-00187>
- Liben, L. S., & Bigler, R. S. (2017). Understanding and undermining the development of gender dichotomies: The legacy of Sandra Lipsitz Bem. *Sex Roles*, 76(9–10), 544–555. <https://doi.org/10.1007/s11199-015-0519-4>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.
- Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). *Single-sex versus coeducational schooling: A systematic review*. American Institutes for Research.
- McGuire, W. J. (1984). Search for the self: Going beyond self-esteem and the reactive self. In R. A. Zucker, J. Aronoff, & A. I. Rabin (Eds.), *Personality and the prediction of behavior* (pp. 73–120). Academic Press.
- McGuire, W. J., McGuire, C. V., & Winton, W. (1979). Effects of household sex composition on the salience of one's gender in the spontaneous self-concept. *Journal of Experimental Social Psychology*, 15(1), 77–90. [https://doi.org/10.1016/0022-1031\(79\)90020-9](https://doi.org/10.1016/0022-1031(79)90020-9)
- McGuire, W. J., & Padawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, 33(6), 743–754. <https://doi.org/10.1037/0022-3514.33.6.743>
- McHale, S. M., Crouter, A. C., & Whiteman, S. D. (2003). The family contexts of gender development in childhood and adolescence. *Social Development*, 12(1), 125–148. <https://doi.org/10.1111/1467-9507.00225>
- Mehta, C. M. (2015). Gender in context: Considering variability in Wood and Eagley's traditions of gender identity. *Sex Roles*, 73(11–12), 490–496. <https://doi.org/10.1007/s11199-015-0535-4>
- 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/10.1016/j.dr.2009.06.001>
- Nagengast, B., Marsh, H. W., & Hau, K. (2013). Effects of single-sex schooling in the final years of high school: A comparison of analysis of covariance and propensity score matching. *Sex Roles*, 69(7–8), 404–422. <https://doi.org/10.1007/s11199-013-0261-8>
- Oberski, D. (2014). Lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, 57(1), 1–27. <https://doi.org/10.18637/jss.v057.i01>
- Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, 120(4), 1013–1034. <https://doi.org/10.1037/pspp0000358>
- Pahlke, E., Bigler, R. S., & Patterson, M. M. (2014). Reasoning about single-sex schooling for girls among students, parents, and teachers. *Sex Roles*, 71(5–8), 261–271. <https://doi.org/10.1007/s11199-014-0410-8>
- Pahlke, E., & Hyde, J. S. (2016). The debate over single-sex schooling. *Child Development Perspectives*, 10(2), 81–86. <https://doi.org/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(4), 1042–1072. <https://doi.org/10.1037/a0035740>
- Palomares, N. A. (2008). Explaining gender-based language use: Effects of gender identity salience on references to emotion and tentative language in intra- and intergroup contexts. *Human Communication Research*, 34(2), 263–286. <https://doi.org/10.1111/j.1468-2958.2008.00321.x>
- Paredes, V. (2022). Mixed but not scrambled: Gender gaps in coed schools with single-sex classrooms. *Journal of Research on Educational Effectiveness*, 15(2), 330–366. <https://doi.org/10.1080/19345747.2021.1966860>
- Patterson, M. M. (2012). Self-perceived gender typicality, gender-typed attributes, and gender stereotype endorsement in elementary-school-aged children. *Sex Roles*, 67(7–8), 422–434. <https://doi.org/10.1007/s11199-012-0184-9>
- Perry, D. G., Pauletti, R. E., & Cooper, P. J. (2019). Gender identity in childhood: A review of the literature. *International Journal of Behavioral Development*, 43(4), 289–304. <https://doi.org/10.1177/0165025418811129>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783. <https://doi.org/10.1037/0022-3514.90.5.751>
- Pishgar, F., Greifer, N., Leyrat, C., & Stuart, E. (2021). Matchthem: Matching and weighting after multiple imputation. *The R Journal*, 13(2), Article 228. <https://doi.org/10.32614/RJ-2021-073>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Riordan, C. (2011). *Single-sex schools: A place to learn*. Rowman & Littlefield.
- Robinson, D. B., Mitton, J., Hadley, G., & Kettley, M. (2021). Single-sex education in the 21st century: A 20-year scoping review of the literature. *Teaching and Teacher Education*, 106, Article 103462. <https://doi.org/10.1016/j.tate.2021.103462>
- Rogers, L. O. (2020). I'm kind of a feminist: Using master narratives to analyze gender identity in middle childhood. *Child Development*, 91(1), 179–196. <https://doi.org/10.1111/cdev.13142>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (1st ed.). John Wiley & Sons. <https://doi.org/10.1002/9780470316696>
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469), 322–331. <https://doi.org/10.1198/016214504000001880>
- Sax, L. (2011, September 26). *Are single-sex schools actually dangerous?* [Web blog]. <https://www.psychologytoday.com/blog/sax-sex/201109>
- Schroeder, K. M., & Liben, L. S. (2021). Felt pressure to conform to cultural gender roles: Correlates and consequences. *Sex Roles*, 84(3–4), 125–138. <https://doi.org/10.1007/s11199-020-01155-9>
- Serbin, L. A., & Sprafkin, C. (1986). The salience of gender and the process of sex typing in three- to seven-year-old children. *Child Development*, 57(5), 1188–1199. <https://doi.org/10.2307/1130442>
- Shutts, K., Kenward, B., Falk, H., Ivegran, A., & Fawcett, C. (2017). Early preschool environments and gender: Effects of gender pedagogy in Sweden. *Journal of Experimental Child Psychology*, 162, 1–17. <https://doi.org/10.1016/j.jecp.2017.04.014>
- 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. <https://doi.org/10.1007/s11199-013-0313-0>
- Signorella, M. L., Frieze, I. H., & Hershey, S. W. (1996). Single-sex versus mixed-sex classes and gender schemata in children and adolescents: A longitudinal comparison. *Psychology of Women Quarterly*, 20(4), 599–607. <https://doi.org/10.1111/j.1471-6402.1996.tb00325.x>
- 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(7–8), 423–441. <https://doi.org/10.1007/s11199-013-0288-x>

- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>
- Sullivan, A., Joshi, H., & Leonard, D. (2010). Single-sex schooling and academic attainment at school and through the life course. *American Educational Research Journal*, 47(1), 6–36. <https://doi.org/10.3102/0002831209350106>
- Sullivan, A., Joshi, H., & Leonard, D. (2012). Single-sex and co-educational secondary schooling: What are the social and family outcomes, in the short and longer term? *Longitudinal and Life Course Studies*, 3(1), 137–156. <https://doi.org/10.14301/llds.v3i1.148>
- Tobin, D. D., Menon, M., Menon, M., Spatta, B. C., Hodges, E. V. E., & Perry, D. G. (2010). The intrapsychics of gender: A model of self-socialization. *Psychological Review*, 117(2), 601–622. <https://doi.org/10.1037/a0018936>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399. <https://doi.org/10.1002/sim.4067>
- Williams, M. W., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences of the United States of America*, 112(17), 5360–5365. <https://doi.org/10.1073/pnas.1418878112>
- Wong, W. I., Shi, S. Y., & Chen, Z. (2018). Students from single-sex schools are more gender-salient and more anxious in mixed-gender situations: Results from high school and college samples. *PLOS ONE*, 13(12), Article e0208707. <https://doi.org/10.1371/journal.pone.0208707>
- Wong, W. I., & VanderLaan, D. P. (2020). Sex differences in early life: A cross-cultural perspective. In F. M. Cheung, & D. F. Halpern (Eds.), *Cambridge international handbook on psychology of women* (pp. 83–95). Cambridge University Press. <https://doi.org/10.1017/9781108561716>
- Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles*, 73(11-12), 461–473. <https://doi.org/10.1007/s11199-015-0480-2>

Received July 25, 2023

Revision received March 24, 2024

Accepted April 1, 2024 ■